

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:55:42 ; Search time 720.588 Seconds  
(without alignments)  
8741.717 Million cell updates/sec

Title: US-10-814-858A-2

Perfect score: 130

Sequence: 1 tcaatctatataggaagttc.....ttacatctagataaacaatg 130

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_hgt:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vi:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	100.0	452	6	E64760
2	130	100.0	452	6	AR427842
3	89	68.5	89	6	E64772
4	89	68.5	89	6	AR427850
C 5	63.6	48.9	9937	12	AY560325
C 6	63.6	48.9	9946	12	AY560326
7	62	47.7	735	12	AY183361
8	62	47.7	827	12	AY192160
C 9	62	47.7	15382	12	AY178047
C 10	62	47.7	16368	12	AY178049
C 11	62	47.7	16520	12	AY178048
12	61	46.9	1006	6	A94910
13	61	46.9	1006	6	BD134819
14	60.4	46.5	1138	6	A59870
15	60.4	46.5	1138	6	A94795
16	60	46.2	12607	12	AY737283
17	58.2	44.8	2200	6	AX451740
18	56.2	43.2	5033	6	I28266
19	52	40.0	275	14	TOTMV6

20	52	40.0	6395	6	AR173320	AR173320 Sequence
21	52	40.0	6395	6	AR271575	AR271575 Sequence
22	52	40.0	6395	6	AX040174	AX040174 Sequence
23	52	40.0	6395	6	AX098414	AX098414 Sequence
24	52	40.0	6395	6	AX194392	AX194392 Sequence
25	52	40.0	6395	14	AF165190	AF165190 Tobacco m
26	52	40.0	6395	14	AF273221	AF273221 Tobacco m
27	52	40.0	6395	14	AF395127	AF395127 Tobacco m
28	52	40.0	6395	14	AF395128	AF395128 Tobacco m
29	52	40.0	6395	14	AF395129	AF395129 Tobacco m
30	52	40.0	6395	14	TWV11933	AJ011933 Tobacco m
31	52	40.0	6395	14	TWVCG	X68110 Tobacco mos
32	52	40.0	6395	14	TOTMV4	V01408 Tobacco mos
33	52	40.0	6425	6	AR173322	AR173322 Sequence
34	52	40.0	6425	6	AX098416	AX098416 Sequence
35	52	40.0	6439	6	AR173321	AR173321 Sequence
36	52	40.0	6439	6	AX098415	AX098415 Sequence
37	52	40.0	6446	6	AR173324	AR173324 Sequence
38	52	40.0	6446	6	AX098418	AX098418 Sequence
39	52	40.0	6475	6	AR173323	AR173323 Sequence
40	52	40.0	6475	6	AX098417	AX098417 Sequence
41	52	40.0	7926	6	BD235479	BD235479 Method fo
42	52	40.0	7926	6	AR224869	AR224869 Sequence
43	52	40.0	7926	6	AR397616	AR397616 Sequence
44	52	40.0	11641	6	AX466963	AX466963 Sequence
45	51.8	39.8	1825	6	AR042908	AR042908 Sequence

## ALIGNMENTS

RESULT 1

E64760

LOCUS  
DEFINITION

E64760 452 bp DNA linear PAT 31-JAN-2002  
Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PN JP 2000139477-A/1

PD 23-MAY-2000

PF 27-APR-1999 JP 1999119227

PR 09-NOV-1998 IN 3322/98

PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,

PI SHIFU KUMAR GUPUTA

CC C12N15/09, C12N5/10, C12N15/00, C12N5/00

PC

FT Key

FT source

FT Location/Qualifiers

FT 1..452

FT /organism='Artificial Sequence'

FT /organism='synthetic construct'

FT /mol\_type='genomic DNA'

FT /db\_xref='taxon:32630'

ORIGIN

Query Match

Best Local Similarity

Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCACATATATAGGAAGTTCAATTCATTTTGAATGACACGTTGTTGTCATTTCTCAACAA 60



TITLE Binary expression vector for use in plants  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 9937)  
AUTHORS Ouwkerk,P.B.F. and Meijer,A.H.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-2004) Molecular Cell Biology, Institute of  
Biology, Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The  
Netherlands

FEATURES  
source  
1. .9937  
/organism="Cloning vector pC1300intB-35SnosEX"  
/mol\_type="other DNA"  
/db\_xref="taxon:267430"  
/notes="derived from pC1300intB (Genbank Accession Number  
AF294977)"  
complement(52. .303)  
/note="3' NOS; terminator signal of nopaline synthase gene"  
complement(322. .347)  
/note="multiple cloning site; SpeI, XhoI, EcoRI, Sall are  
unique"  
complement(425. .789)  
/note="CamV 35S promoter"  
1066. .1091  
/note="right border T-DNA repeat"  
complement(2132. .3132)  
/note="STa region from pVS1 plasmid"  
complement(3725. .4725)  
/note="pVS1-REP; replication origin from pVS1"  
complement(5135. .5395)  
/note="bom site from pBR322"  
complement(5535. .5815)  
/note="pBR322 origin of replication"  
/genes="aadA"  
complement(6106. .6900)  
complement(6106. .6900)  
/genes="aadA"  
/note="AadA; kanamycin resistance gene from pIG121Hm"  
/codon\_start=1  
/transl\_table=11  
/product="aminoglycoside transferase"  
/protein\_id="AA573283.1"  
/db\_xref="GI:45645332"  
/translation="MKMRISPELKKLIEKYRCVKDTGMSPAKVYKLVGENENLYLK  
MTDSRYGTTYDVERGKMLWLEGLPVPKVLHFERHDSNLLMSEADGVLCSBEY  
EDQSPKIELYAEICIRLPHSIDCPVTNSLDRSLAELDYLLANDLADYDVCENWE  
EDFPKDPRELYDLFETKEPEELVFESHGDLGDSNIFVKDGKVSGLFIDLGRGRADKW  
YDIAPCVRSIREDIGEEQVELFDFLLGKPKDWEKIKYLLIDELF"  
7325. .7350  
/note="left border repeat from C58 T-DNA"  
complement(7417. .7936)  
/note="CamV 35S promoter"  
complement(7417. .7636)  
/note="CamV polyA signal"  
complement(7652. .8867)  
/genes="hptII"  
complement(join(7652. .8338,8529. .8867))  
/genes="hptII"  
/note="hptII"  
/codon\_start=1  
/transl\_table=11  
/product="hygromycin phosphotransferase"  
/protein\_id="AA573284.1"  
/db\_xref="GI:45645333"  
/translation="MKKPELTATSVKFLIEKFDSDVSLMQLSEGESRAFPEDVGGRR  
GYIVRVNSCAGGYKDYVIRHPASALPIPEVLDDIGFESLTYCISRAAGVTLQD  
LPETELPAVLQVPAEMDAIAADLSQTSFGPGFGQIGQYTTWRDFTCALADPHVY  
HMQTVMDTYSVAQALDELMLAEDCEPRLVHADFGSNVLTNDNGRITAVIDWS  
EAMFGDSQYEVANIFFWRPWLACMEQOTRYFERRHPELAGSPRLRAYMLRIGLDOLYO  
SLVDGNFDAAWAQSCDAIVRSAGTVGRTQIARRSAAWTDGCGVEVLADSGNRPRS  
TPRPAKK"  
complement(7652. .8338)  
/genes="hptII"  
/number=2

terminator  
misc\_feature  
promoter  
misc\_feature  
misc\_feature  
rep\_origin  
misc\_feature  
rep\_origin  
gene  
CDS

introns  
complement(8339. .8534)  
/genes="hptII"  
/note="derived from Ricinus communis catalase-1 intron  
presented in Genbank Accession Number D21161; confers  
higher resistance towards hygromycin in transgenic rice  
and allows easier killing of Agrobacterium tumefaciens  
during regeneration of transgenic plants"  
complement(8529. .8867)  
/genes="hptII"  
/number=1

exon  
Query Match 48.9%; Score 63.6; DB 12; Length 9937;  
Best Local Similarity 87.7%; Pred. No. 5.7e-05;  
Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;  
QY 12 AGGAAGTTTCATTTCATTTCGATGGACA--CGTGTTCATTTCTCAACAATTACCAACA 69  
|||||  
Db 451 AGGAAGTTTCATTTCATTTCGATGGAGGACAGGCTTCTTGAGATCCTTCAACAATTACCAACA 392  
|||||  
QY 70 ACAACAACAACAACAACAATATATACAAATTACTATTATTACAAATTACA 115  
|||||  
Db 391 ACAACAACAACAACAACAAT--TACAATTACTATTATTACAAATTACA 348  
|||||

RESULT 6  
AY560326/c 9946 bp DNA circular SYN 28-MAR-2004  
LOCUS Cloning vector pC1300intB-35SnosBK, complete sequence.  
DEFINITION AY560326  
ACCESSION AY560326  
VERSION AY560326.1 GI:45645334  
KEYWORDS  
SOURCE Cloning vector pC1300intB-35SnosBK  
ORGANISM Cloning vector pC1300intB-35SnosBK  
REFERENCE 1 (bases 1 to 9946)  
AUTHORS Ouwkerk,P.B.F. and Meijer,A.H.  
TITLE Binary expression vector for use in plants  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 9946)  
AUTHORS Ouwkerk,P.B.F. and Meijer,A.H.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-2004) Molecular Cell Biology, Institute of  
Biology, Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The  
Netherlands

FEATURES  
source  
1. .9946  
/organism="Cloning vector pC1300intB-35SnosBK"  
/mol\_type="other DNA"  
/db\_xref="taxon:267431"  
complement(48. .239)  
/note="3' NOS; terminator signal of nopaline synthase gene"  
complement(318. .356)  
/note="multiple cloning site; Sall, EcoRI, XhoI, BglII,  
KpnI, SpeI are unique"  
complement(434. .798)  
/note="CamV 35S promoter"  
1075. .1100  
/note="right border T-DNA repeat"  
complement(2141. .3141)  
/note="STa region from pVS1 plasmid"  
complement(3734. .4734)  
/note="pVS1-REP; replication origin from pVS1"  
complement(5144. .5404)  
/note="bom site from pBR322"  
complement(5544. .5824)  
/note="pBR322 origin of replication"  
complement(6115. .6909)  
/genes="aadA"  
complement(6115. .6909)  
/genes="aadA"  
/note="AadA; kanamycin resistance gene from pIG121Hm"  
/codon\_start=1

terminator  
misc\_feature  
promoter  
misc\_feature  
misc\_feature  
rep\_origin  
misc\_feature  
rep\_origin  
gene  
CDS  
exon

```
/transl_table=11
/product="aminoglycoside transferase"
/protein_id="AAS73285.1"
/db_xref="GI:45645335"
/translations="MAKMRISPEKKLIEKRYCVKDTQEGMSPAKVYKLVGENENLYLK
MTDSRYKGTYYDVREKDWMLWEGKLPVPKVLHFERHDGWSNLLMSEADGVLCSEY
EDQSPEKILAEICIRLPHSIDSDCPYTNLSRLAEILDYLLNNDLADVDCEWNE
EDTPFDKPRELYDFLTERPEELVFSHGLDGSNIFVKDGKVSFGFIDLGRGRADKW
YDIAFCVRSIREDEEQVELFDLLGLIKPWEKIKYILIDELF"
/misc_feature
complement(7334..7359)
/notes="left border repeat from C58 T-DNA"
/polyA_signal
complement(7426..7645)
/notes="CaMV polyA signal"
gene
complement(7661..8876)
/genes="hptII"
CDS
complement(join(7661..8347,8538..8876))
/genes="hptII"
/notes="hptII"
/codon_start=1
/transl_table=11
/product="hygromycin phosphotransferase"
/protein_id="AAS73286.1"
/db_xref="GI:45645336"
/translations="MKKPELTATSVKFLIEKFDVSVDLMQLSEBSRAFPDFVGGQ
GVYLRVNSCADGPYKDYVYVRHPASAAPIPEVLIDIGFSESITVCISRAQGVTLQD
LPETELPAVLQPVAEAMDAIAADLSQTSQGFPGQIGQYTHRDFICADAPHYV
HWQVMDTTSVASVAQLDELMLWAECDCEVRLHVLHADFGSNVLLDNGRITAVIDWS
EAMFGDSQYEVANIFFWRPWLACMEQOTYFERRHPHELAGSPRLRAYMLRIGLDQLYQ
SLVDGNFDDAANAQGRCDATVRSAGTGTGTQIARRSAAVWTDGCVVLADSGNRRPS
TRPRAKK"
/exon
complement(7661..8347)
/genes="hptII"
/intron
complement(8348..8537)
/genes="hptII"
/notes="derived from Ricinus communis catalase-1 intron
presented in Genbank Accession Number D21161; confers
higher resistance towards hygromycin in transgenic rice
and allows easier killing of Agrobacterium tumefaciens
during regeneration of transgenic plants"
/exon
complement(8538..8876)
/genes="hptII"
/promoter
complement(8923..9703)
/notes="35S promoter from CaMV"
```

```
ORIGIN
Query Match 48.9%; Score 63.6; DB 12; Length 9946;
Best Local Similarity 87.7%; Pred. No. 5.7e-05;
Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACA--CGTGTGTGTCATTTCTCAACAATTACCAACA 69
Db 460 AGGAAGTTCATTTTCATTTGGAATGGACAGGCTCTTGGAGATCTCTCAACAATTACCAACA 401
Qy 70 ACAACAACAACAACAACAACAATTACTATTACTATTACTATTACTATTACTATTACTATTACTA 115
Db 400 ACAACAACAACAACAACAACAACAT--TACAATTACTATTACTATTACTATTACTATTACTA 357

RESULT 7
AY183361
LOCUS
DEFINITION
Synthetic construct duplicated CamV 35/TMV omega-prime leader
sequence fusion promoter, complete sequence.
ACCESSION
AY183361
VERSION
AY183361.1 GI:28193426
KEYWORDS
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 735)
AUTHORS
Yao,Q., Peng,R. and Xiong,A.
```

```
Direct Submission
Submitted (19-NOV-2002) Agro-Biotechnology Research Center,
Shanghai Academy of Agriculture Science, Beidi Road 2901, Shanghai
201106, China
FEATURES
source
1..735
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
promoter
1..735
/notes="duplicated CamV 35/TMV omega-prime leader sequence
fusion promoter"
ORIGIN
Query Match 47.7%; Score 62; DB 12; Length 735;
Best Local Similarity 88.5%; Pred. No. 0.00019;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

Qy 12 AGGAAGTTCATTTTCATTTGGAATGGACACACGTGTGTGTCATTTCTCAACAATTACCAACAAC 71
Db 639 AGGAAGTTCATTTTCATTTGGAATGGACACACGTAT-----TTTACACAATTACCAACAAC 693
Qy 72 AACAAACAACAACAACAATTACTATTACTATTACTATTACTATTACTATTACTATTACTATTACTA 115
Db 694 AACAAACAACAACAACAAT--TACAATTACTATTACTATTACTATTACTATTACTATTACTA 735

RESULT 8
AY192160
LOCUS
DEFINITION
Synthetic construct multi-copy enhancer promoter element, complete
sequence.
ACCESSION
AY192160
VERSION
AY192160.1 GI:28569528
KEYWORDS
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 827)
AUTHORS
Yao,Q., Peng,R. and Xiong,A.
TITLE
An artificial promoter containing many copies of enhancers can
increase phytase expression in transgenic rape
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 827)
AUTHORS
Yao,Q., Peng,R. and Xiong,A.
Direct Submission
Submitted (07-DEC-2002) Biotechnology Research Center, Shanghai
Academy of Agriculture Science, Beidi Road 2901, Shanghai, Shanghai
201106, China
FEATURES
source
1..827
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
promoter
1..827
/notes="multi-copy enhancer promoter element"
misc_feature
1..90
/notes="contains 3 copies of 30bp element from glycine
promoter"
91..160
/notes="contains 3 copies of 22bp W1 seed-specific
cis-element from pea lectin promoter"
misc_feature
161..376
/notes="contains 3 copies of 68bp UAS from phaseolin
promoter"
501..827
/notes="contains CaMV mini promoter (-60) and TMV omega
enhancer"
ORIGIN
Query Match 47.7%; Score 62; DB 12; Length 827;
Best Local Similarity 88.5%; Pred. No. 0.00019;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;
```



```
Oy 12 AGGAAGTTCATTTCATTGGAATGGACACGTGTGTCATTCTTCAACAATTACCAACAAC 71
|||||
Db 731 AGGAAGTTCATTTCATTGGAAGGACACGTAT-----TTTACACAATTACCAACAAC 785
|||||

Oy 72 AACAAACAAACAAACAAATATACAAATTACTATTACAAATTACA 115
|||||
Db 786 AACAAACAAACAAACAAACAT--TACAATTACTATTACAAATTACA 827
|||||

RESULT 9
AY178047/c 15382 bp DNA circular SYN 06-APR-2003
LOCUS Expression vector pYPX143, complete sequence.
DEFINITION Expression vector pYPX143
ACCESSION AY178047
VERSION AY178047.1 GI:29568870
KEYWORDS Expression vector pYPX143
SOURCE Expression vector pYPX143
ORGANISM other sequences; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 15382)
AUTHORS Yao, Q., Peng, R. and Xiong, A.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Shanghai Academy of Agricultural Sciences,
Agro-Biotechnology Research Center, 2901 Beidi Rd, Shanghai 201106,
China
COMMENT NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source
1. .15382
/organism="Expression vector pYPX143"
/mol_type="genomic DNA"
/db_xref="taxon:218199"
/lab_host="Agrobacterium tumefaciens"
complement(5139..5798)
/notes="synthetic sequence derived from castorbean"
/codon_start=1
/product="catalase"
/protein_id="AAO84026.1"
/db_xref="GI:29568871"
/translation="MEKKTIGTVDISQHRKEHPEAFQSVQAQCTYNTQVOLDITAP
LKTAKGKHKFPATFHLARLNHAPFPMAMKOGELVWDSVHPCTYTFHEQTET
SLWSYHDFDQFLHYSDVACGENLAYPPKGFENMFVSANPWSYFTSFDLNV
ANNNDFFAPVFTMGKYTQGDYKMLPLAQVHHAUCDFHVGMLNELQYCDWQGG
A"
misc_feature
complement(529..7504)
/notes="similar to aminoglycoside-3'-O-phosphotransferase"
CDS
complement(8421..10232)
/codon_start=1
/product="Gus"
/protein_id="AAO84027.1"
/db_xref="GI:29568872"
/translation="MLRPVETPTREIKKLDGLWAFSLDRENCIGDORWWSALQBSRA
IAPGFSNQFADADIRNAGNVYQREVFIPKGWAGQRIVLRFDAVTHYGVWNNQ
EYMEHGGYTPFEADPTVIAGKSVRIITCVNNELNMQTIIPGMVITDENGKKQSY
FHDFFNYAGIHRSMVLYTTPNTWDDITVTHVAQDNHASVDWQVANGDYSVELRD
AQOQVATTCGTSGLQVNNPHLWPGEGVLYELCVTAQKOTCEDIYPLRVGIRSVAV
KGEQFLNKHGIPYFTGTFGRHEDALGKGFNDVNLVHDHMDWJANSYRTSHYPA
SEMLDWDHSHGIVVIDETAAGVNLGLIGFEGAGNPKELIYSEEVANGSTQQAHLQAI
KELIARDKNHPSVMMWSIANEPDTRPGQAREYFAPLAETAKLDLPTPTTCVNNMFC
AHTDITLSDFLCLNRYGTGWYQSGDLEAKVLEKELLAWQKLHQPILIIITEXGVD
TLAGLSMTYTDWMSYEQCAWLMDMYHRVDFRVSADVGEQVWNPADFPATFSQILRVGGN
KKGIFTRDRKPKSAEFLQKRWTKMNFGEKPKQGGKQ"
join(12923..12937,13128..14975)
/codon_start=1
/product="his tagged GusA"
/protein_id="AAO84028.1"
/db_xref="GI:29568873"
/translation="WVDLENRLRVETPTREIKKLDGLWAFSLDRENCIGDORWWE
SALQBSRAIAPGFSNQFADADIRNAGNVYQREVFIPKGWAGQRIVLRFDAVTHY
GKWNQNVMEHGGYTPFEADPTVIAGKSVRIITCVNNELNMQTIIPGMVITDE
NGKKQSYFHDFFNYAGIHRSMVLYTTPNTWDDITVTHVAQDNCHASVDWQVANG
DYSVELRDADQVATTCGTSGLTQVNNPHLWPGEGVLYELVTAQKOTCEDIYPLR
```

```
VGIRSVAVKGOQFLINHKPFYFTGFRHEDADLRGKGFNDVNLVHDHMDWJANSY
RTSHYPYAEEMLDWADHGIIVVIDETAAGVNLGLIGFEGAGNPKELIYSEEVANGST
QQAHLQAIKELIARDKNHPSVMMWSIANEPDTRPGQAREYFAPLAETAKLDLPTPTIT
CVNNMFCADHTDITLSDFLCLNRYGTGWYQSGDLEAKVLEKELLAWQKLHQPILIIIT
EYGVDTLAGLSMTYTDWMSYEQCAWLMDMYHRVDFRVSADVGEQVWNPADFPATFSQ
GILRVGGNKKGIFTRDRKPKSAEFLQKRWTKMNFGEKPKQGGKQASHHHHHV"

Query Match 47.7%; Score 62; DB 12; Length 15382;
Best Local Similarity 88.5%; Pred. No. 0.00012;
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

Oy 12 AGGAAGTTCATTTCATTGGAATGGACACGTGTGTCATTCTTCAACAATTACCAACAAC 71
|||||
Db 10335 AGGAAGTTCATTTCATTGGAAGGACACGTAT-----TTTACACAATTACCAACAAC 10281
|||||

Oy 72 AACAAACAAACAAACAAATATACAAATTACTATTACAAATTACA 115
|||||
Db 10280 AACAAACAAACAAACAAACAT--TACAATTACTATTACAAATTACA 10239
|||||

RESULT 10
AY178049/c 16368 bp DNA circular SYN 06-APR-2003
LOCUS Expression vector pYPX245, complete sequence.
DEFINITION Expression vector pYPX245
ACCESSION AY178049
VERSION AY178049.1 GI:29568877
KEYWORDS Expression vector pYPX245
SOURCE Expression vector pYPX245
ORGANISM other sequences; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 16368)
AUTHORS Yao, Q., Peng, R. and Xiong, A.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Shanghai Academy of Agricultural Sciences,
Agro-Biotechnology Research Center, 2901 Beidi Rd, Shanghai 201106,
China
COMMENT NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source
1. .16368
/organism="Expression vector pYPX245"
/mol_type="genomic DNA"
/db_xref="taxon:218201"
/lab_host="Agrobacterium tumefaciens"
complement(4955..5749)
/gene="aadA"
complement(4955..5749)
/gene="aadA"
/function="kanamycin resistance"
/codon_start=1
/product="aminoglycoside phosphotransferase"
/protein_id="AAO84031.1"
/db_xref="GI:29568878"
/translation="MAKMRISPELKKLIEKRVCKDTEGMSPAKYVKLVGENENLYLK
MTDSYKGTTYDVEREKDMLWLEKLPVKLPHERHGDWNSLSEADGVLCSSEY
EDQSPKIIELYAFCLRIELHSIDISDCPYTNLSLRALAEYLNLNDLADVDCEWNE
BDTQPEKRLDYDLTKPEELVPSHGLDGSNIYFKDGKVSQFIDLGRGRADKW
YDIAFCVRSIRREDIGEEQYVELFDLLGIGKPDWEKIKYILLDELFP"
misc_feature
complement(6377..7352)
/notes="similar to aminoglycoside-3'-O-phosphotransferase"
gene
complement(9437..11248)
/gene="gus"
complement(9437..11248)
/gene="gus"
CDS
complement(9437..11248)
/gene="gus"
/codon_start=1
/product="Gus"
/protein_id="AAO84032.1"
/db_xref="GI:29568879"
/translation="MLRPVETPTREIKKLDGLWAFSLDRENCIGDORWWSALQBSRA
IAPGFSNQFADADIRNAGNVYQREVFIPKGWAGQRIVLRFDAVTHYGVWNNQ
EYMEHGGYTPFEADPTVIAGKSVRIITCVNNELNMQTIIPGMVITDENGKKQSY
FHDFFNYAGIHRSMVLYTTPNTWDDITVTHVAQDNCHASVDWQVANGDYSVELRD
```

```
ADQOVATCGTSGTLOVVNPHLMQPGEGYELCVAKTSQTECDIYPLRGIRSVAV
KGEQFLINHKPFYFTGFRHEDADLRGKGFDMVMDHMDWIGANSYRTSHYPIYA
EEMLDWADHSHGIVVIDETAAGVFNLSLGI GFAGNKPKELYSEEA VNGETQOAHLOAI
KELIARDKHSHVVMWSIANBPDRPQAGREYFAPLAETRLKLDPTPTTCVNMVFCD
AHTDITSDIFDVLCLNRYGYWQSGDLETAEKVLEKELLAWQEKLHQPILIIITEYVD
TLAHLMSMTDMWSEYOCALMDMTHRVFDRVSADVGEQVWNFADPATSQGLIRVGGN
KKGIFTRDRKPKSAAPLLQKRWTMGNFGEKPOQGGKQ"
misc_feature
14114..15961
/ note="similar to his tagged Gusa"
```

ORIGIN

Query Match 47.7%; Score 62; DB 12; Length 16368;  
Best Local Similarity 88.5%; Pred. No. 0.00012;  
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71  
|||||  
Db 11351 AGGAAGTTCATTTCATTTGGAATGGACACGCTAT-----TTTACACAATTACCAACAAC 11297  
|||||

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 115  
|||||  
Db 11296 AACAAACAACAACAACAT--TACAATTACTATTACAAATTACA 11255  
|||||

RESULT 11  
AY178048/c 16520 bp DNA circular SYN 06-APR-2003  
LOCUS  
DEFINITION Expression vector pYPX145, complete sequence.  
ACCESSION AY178048  
VERSION AY178048.1 GI:29568874  
KEYWORDS Expression vector pYPX145  
SOURCE Expression vector pYPX145  
ORGANISM other sequences; artificial sequences; vectors.  
REFERENCE 1 (bases 1 to 16520)  
AUTHORS Yao Q., Peng, R. and Xiong, A.  
TITLE Direct Submission  
JOURNAL Submitted (12-NOV-2002) Shanghai Academy of Agricultural Sciences,  
Agro-Biotechnology Research Center, 2901 Beidi Rd, Shanghai 201106,  
China  
COMMENT NCBI staff are still waiting for submitters to provide appropriate  
coding region information.  
FEATURES  
source  
1..16520  
/organism="Expression vector pYPX145"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:218200"  
/lab\_host="Agrobacterium tumefaciens"  
/gene="cat"  
complement(5139..5798)  
complement(5139..5798)  
/gene="cat"  
/codon\_start=1  
/product="chloramphenicol acetyl transferase"  
/protein\_id="AA084029.1"  
/db\_xref="GI:29568875"  
/translation="MEKLTGYTVDISOMHKEHFEAPQSVAQCTYNTQVOLDITAF  
LTKVKNKHGFYAFTHILARLNAHPEFRAMKDGELVINDSVHPCTYVPEQTEF  
SSLWSEYHDDFQFLHYISQDVACGENLAYPPKGFIEFMFVSANPWVSTSPFLNV  
ANMNDFFAPVFTMGKYYTQGDVKVLMPLAIQVHHVAVCDGFHVGRLNELQYCDQWQGG  
A"

misc\_feature  
complement(6529..7504)  
/note="similar to aminoglycoside-3'-O-phosphotransferase"  
complement(9589..11400)  
/gene="gus"  
complement(9589..11400)  
/gene="gus"  
/codon\_start=1  
/product="Gus"  
/protein\_id="AA084030.1"  
/db\_xref="GI:29568876"  
/translation="MLRPVETPTREIKKLDGLWAFSLDRENGCIDQRWWSALQESRA  
IAPGSFNDQPADADIRPNVAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGVWYNNQ

```
EVMHEQGYTPEADYTPYVLAGKSVRLTVCNNELNHQTLPQGVITDENGKKKOSY
FHDFFNYAGIHRSVNMLYTTPTNTWDDIIVTHVAQDCNHSVDWQVANGDVSVLELRD
ADQOVATCGTSGTLOVVNPHLMQPGEGYELCVAKTSQTECDIYPLRGIRSVAV
KGEQFLINHKPFYFTGFRHEDADLRGKGFDMVMDHMDWIGANSYRTSHYPIYA
EEMLDWADHSHGIVVIDETAAGVFNLSLGI GFAGNKPKELYSEEA VNGETQOAHLOAI
KELIARDKHSHVVMWSIANBPDRPQAGREYFAPLAETRLKLDPTPTTCVNMVFCD
AHTDITSDIFDVLCLNRYGYWQSGDLETAEKVLEKELLAWQEKLHQPILIIITEYVD
TLAHLMSMTDMWSEYOCALMDMTHRVFDRVSADVGEQVWNFADPATSQGLIRVGGN
KKGIFTRDRKPKSAAPLLQKRWTMGNFGEKPOQGGKQ"
misc_feature
14251..16113
/ note="similar to his tagged Gusa"
```

ORIGIN

Query Match 47.7%; Score 62; DB 12; Length 16520;  
Best Local Similarity 88.5%; Pred. No. 0.00012;  
Matches 92; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71  
|||||  
Db 11503 AGGAAGTTCATTTCATTTGGAATGGACACGCTAT-----TTTACACAATTACCAACAAC 11449  
|||||

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 115  
|||||  
Db 11448 AACAAACAACAACAACAT--TACAATTACTATTACAAATTACA 11407  
|||||

RESULT 12  
A94910 1006 bp DNA linear PAT 26-JAN-2000  
LOCUS  
DEFINITION Sequence 22 from Patent WO9931258.  
ACCESSION A94910  
VERSION A94910.1 GI:6779107  
KEYWORDS  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1006)  
AUTHORS Stuiver, M.H. and Siboltz, F.H.  
TITLE CONSTITUTIVE PLANT PROMOTERS  
JOURNAL Patent: WO 9931258-A 22 24-JUN-1999;  
STUIVER MAARTEN HENDRIK (NL); MOGEN INT (NL)  
FEATURES  
source  
1..1006  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

ORIGIN

Query Match 46.9%; Score 61; DB 6; Length 1006;  
Best Local Similarity 88.3%; Pred. No. 0.00031;  
Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAATGGACACGCTGTGTCATTTCTCAACAATTACCAACAAC 71  
|||||  
Db 906 AGGAAGTTCATTTCATTTGGAATGGACACGCTAT-----TTTACACAATTACCAACAAC 960  
|||||

QY 72 AACAAACAACAACAATATACAAATTACTATTACAAATTACA 114  
|||||  
Db 961 AACAAACAACAACAACAT--TACAATTACTATTACAAATTAC 1001  
|||||

RESULT 13  
BD134819 1006 bp DNA linear PAT 18-SEP-2002  
LOCUS  
DEFINITION Constitutive plant promoter.  
ACCESSION BD134819  
VERSION BD134819.1 GI:23229764  
KEYWORDS JP 2002508186-A/22.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1006)  
AUTHORS Stuiver, M.H. and Siboltz, F.H.

```
TITLE      Constitutive plant promoter
JOURNAL    Patent: JP 2002508186-A 22-19-MAR-2002;
COMMENT    MOGEN INTERNATIONAL NV
           OS Unidentified
           PN JP 2002508186-A/22
           PD 19-MAR-2002 JP 2000539156
           PF 10-DEC-1998 JP 2000539156
           PR 12-DEC-1997 EP 97203912.7
           PI MAARTEN HENDRIK STUIVER, FLOOR HENDRIC SIBORTZ PC
           C12N15/09, C12N15/00
           CC Strandedness: Double;
           CC Topology: Linear;
           CC Constitutive plant promoter
           FH Key Location/Qualifiers
           FT source 1..1006
           FT Location/Qualifiers
           source 1..1006
           /organism="unidentified"
           /mol_type="genomic DNA"
           /db_xref="taxon:32644"

ORIGIN
Query Match 46.9%; Score 61; DB 6; Length 1006;
Best Local Similarity 88.3%; Pred. No. 0.00031;
Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71
Db 906 AGGAAGTTCATTTCATTGGATGGACACGTAT-----TTTACAACAATTACCAACAAC 960

QY 72 AACAAACAACAACAACATTATACAAATTACTATTACAATTAC 114
Db 961 AACAAACAACAACAACAT--TACAATTACTATTACAATTAC 1001

RESULT 14
A59870
LOCUS      1138 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 2 from Patent WO9706269.
ACCESSION A59870
VERSION A59870.1 GI:3715061
KEYWORDS   unidentified
SOURCE     unidentified
ORGANISM   unclassified.

REFERENCE 1
AUTHORS    Jepson, I.
TITLE      INDUCIBLE HERBICIDE RESISTANCE
JOURNAL    Patent: WO 9706269-A 2 20-FEB-1997;
           ZENECA LTD (GB)
FEATURES    Location/Qualifiers
           source 1..1138
           /organism="unidentified"
           /mol_type="unassigned DNA"
           /isolate="PLASMID MJB1"
           /db_xref="taxon:32644"

ORIGIN
Query Match 46.5%; Score 60.4; DB 6; Length 1138;
Best Local Similarity 85.8%; Pred. No. 0.00041;
Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71
Db 743 AGGAAGTTCATTTCATTGGAGAGG--ACCTCGAGTATTTTACAACAATTACCAACAAC 800

QY 72 AACAAACAACAACAACATTATACAAATTACTATTACAATTACATC 117
Db 801 AACAAACAACAACAACAT--TACAATTACTATTACAATTACACC 844

RESULT 15
A94795
LOCUS      1138 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 39 from Patent WO9932630.
ACCESSION A94795
VERSION A94795.1 GI:6779045
KEYWORDS   unidentified
SOURCE     unidentified
ORGANISM   unclassified.

REFERENCE 1 (bases 1 to 1138)
AUTHORS    Jones, P.G. and Holt, D.C.
TITLE      HERBICIDE BINDING PROTEINS AND TRANSGENIC PLANTS CONTAINING THEM
JOURNAL    Patent: WO 9932630-A 39 01-JUL-1999;
           JONES PAUL GLYN (GB); ZENECA LTD (GB)
FEATURES    Location/Qualifiers
           source 1..1138
           /organism="unidentified"
           /mol_type="unassigned DNA"
           /db_xref="taxon:32644"
           /clone="NUCLEOTIDE SEQUENCE FROM THE ECOR1 TO HIND III
           RESTRICTION SITES IN PMJB1."

ORIGIN
Query Match 46.5%; Score 60.4; DB 6; Length 1138;
Best Local Similarity 85.8%; Pred. No. 0.00041;
Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTGGATGGACACGTGTGTCATTCTCAACAATTACCAACAAC 71
Db 743 AGGAAGTTCATTTCATTGGAGAGG--ACCTCGAGTATTTTACAACAATTACCAACAAC 800

QY 72 AACAAACAACAACAACATTATACAAATTACTATTACAATTACATC 117
Db 801 AACAAACAACAACAACAT--TACAATTACTATTACAATTACACC 844

Search completed: July 3, 2005, 03:19:09
Job time : 724.588 secs
```

the huge black (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 02:18:08 ; Search time 140.471 Seconds  
(without alignments)  
3634.343 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 Gtcgaccatcattgaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/prodata/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	310.4	99.5	452	4	US-09-263-692A-1
2	46	14.7	46	4	US-09-263-692A-5
3	41.6	13.3	7218	1	US-08-232-463-14
4	36.8	11.8	254778	4	US-09-949-016-12417
5	36.4	11.7	289	3	US-09-007-005-17
6	36.4	11.7	289	3	US-09-244-796-17
7	36.4	11.7	289	3	US-09-244-796-3
8	33.8	10.8	22294	4	US-09-949-016-14020
9	33.6	10.8	248	3	US-09-007-005-32
10	33.6	10.8	248	3	US-09-244-796-32
11	33.6	10.8	277	3	US-09-007-005-3
12	33.6	10.8	277	3	US-09-244-796-3
13	33.4	10.7	601	4	US-09-949-016-78043
14	32.4	10.4	1096	3	US-08-858-207A-136
15	32.4	10.4	11303	3	US-08-861-527-115
16	32.4	10.4	68702	4	US-09-949-016-16328
17	32.4	10.4	102409	4	US-09-949-016-15148
18	32.2	10.3	3270	4	US-09-248-796A-4969
19	31.6	10.1	1958	4	US-09-270-767-10139
20	31.6	10.1	2526	3	US-09-202-712-1
21	31.6	10.1	84495	3	US-09-797-906-3
22	31.2	10.0	491	4	US-09-270-767-7112
23	31.2	10.0	491	4	US-09-270-767-22394
24	31	9.9	601	4	US-09-949-016-164450
25	31	9.9	44715	4	US-09-949-016-16353
26	30.8	9.9	1980	4	US-09-583-110-1382
27	30.8	9.9	1986	4	US-09-107-433-1097

28	30.6	9.8	5053	3	US-08-961-527-187	Sequence 187, Appl
29	30.6	9.8	74962	4	US-09-685-853A-3	Sequence 3, Appl1
30	30.4	9.7	207	4	US-09-248-796A-6311	Sequence 6311, Ap
31	30.4	9.7	89689	4	US-09-949-016-13089	Sequence 13089, A
32	30.4	9.7	153866	4	US-09-949-016-16319	Sequence 16319, A
33	30.2	9.7	477	4	US-09-669-751-206	Sequence 206, App
34	30.2	9.7	1069	4	US-09-270-767-13807	Sequence 13807, A
35	30.2	9.7	177669	4	US-09-949-016-13713	Sequence 13713, A
36	30	9.6	1214	4	US-09-506-286B-15	Sequence 15, Appl
37	30	9.6	1214	4	US-09-506-286B-18	Sequence 18, Appl
38	30	9.6	1214	4	US-09-762-861B-15	Sequence 15, Appl
39	30	9.6	1214	4	US-09-762-861B-18	Sequence 18, Appl
40	30	9.6	1214	4	US-10-065-133A-15	Sequence 15, Appl
41	30	9.6	1214	4	US-10-065-133A-18	Sequence 18, Appl
42	30	9.6	1214	4	US-10-434-811A-15	Sequence 15, Appl
43	30	9.6	1214	4	US-10-434-811A-18	Sequence 18, Appl
44	30	9.6	1241	4	US-09-506-286B-13	Sequence 13, Appl
45	30	9.6	1241	4	US-09-506-286B-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-263-692A-1

; Sequence 1, Application US/09263692A

; Patent No. 6639065

; GENERAL INFORMATION:

; APPLICANT: Council of Scientific and Industrial Research

; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expr

; TITLE OF INVENTION: transgenes and a method for its synthesis

; FILE REFERENCE: Q52511

; CURRENT APPLICATION NUMBER: US/09/263,692A

; CURRENT FILING DATE: 1999-03-05

; PRIOR APPLICATION NUMBER: 3322/Del/98

; PRIOR FILING DATE: 1998-11-09

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 452

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: synthetic DNA promoter sequence

US-09-263-692A-1

Query Match 99.5%; Score 310.4; DB 4; Length 452;

Best Local Similarity 99.7%; Pred. No. 1.7e-94;

Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACCATCTTGAAGGGCTCGGTAATACCATTTGGGAAAAGTTGGTAATACGGA 60

Db 1 GTCGACCATCTTGAAGGGCTCGGTAATACCATTTGGGAAAAGTTGGTAATACGGA 60

QY 61 AAAAGAAGATTCATCATCCAGAAAAGTGTGAAAAAGTTTGCATTGCGTGGAAAAAGTTT 120

Db 61 AAAAGAAGATTCATCATCCAGAAAAGTGTGAAAAAGTTTGCATTGCGTGGAAAAAGTTT 120

QY 121 CGATCTGACCATCTCTAGATCGTGAAAAGTTTCAGTTAGCGCTTACGTACATATGTGG 180

Db 121 CGATCTGACCATCTCTAGATCGTGAAAAGTTTCAGTTAGCGCTTACGTACATATGTGG 180

QY 181 ATTGTGGAAAAGAGACGGAGCATCGTGGAAAAGAGCTTGTACGCTGTACGCTGA 240

Db 181 ATTGTGGAAAAGAGACGGAGCATCGTGGAAAAGAGCTTGTACGCTGTACGCTGA 240

QY 241 CGATAGATAGATACACGTGCGTCACCGCTCCACTTTGACGCCAATTGACGCCAATGACGCCA 300

Db 241 CGATAGATAGATACACGTGCGTCACCGCTCCACTTTGACGCCAATTGACGCCAATGACGCCA 300

QY 301 CTTGACGCTACT 312

Db 301 CTTGACGCTACT 312

## RESULT 2

US-09-263-692A-5  
; Sequence 5, Application US/09263692A  
; Patent No. 6639065  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expression of transgenes and a method for its synthesis  
; FILE REFERENCE: 052511  
; CURRENT APPLICATION NUMBER: US/09/263,692A  
; PRIOR FILING DATE: 1999-03-05  
; PRIOR APPLICATION NUMBER: 3322/Del/98  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 46  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: minimal domain (b)  
US-09-263-692A-5

Query Match 14.7%; Score 46; DB 4; Length 46;  
Best Local Similarity 100.0%; Pred. No. 6.6e-06;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 267 CCACCTTGACGCAACAATTGACGCAATGACGCCCACTTGACGCTACT 312  
Db 1 CCACCTTGACGCAACAATTGACGCAATGACGCCCACTTGACGCTACT 46

## RESULT 3

US-09-263-692A-6  
; Sequence 6, Application US/09263692A  
; Patent No. 6639065  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A chemically synthesized artificial promoter for high level expression of transgenes and a method for its synthesis  
; FILE REFERENCE: 052511  
; CURRENT APPLICATION NUMBER: US/09/263,692A  
; PRIOR FILING DATE: 1999-03-05  
; PRIOR APPLICATION NUMBER: 3322/Del/98  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 46  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: domain I  
US-09-263-692A-6

Query Match 14.7%; Score 46; DB 4; Length 46;  
Best Local Similarity 100.0%; Pred. No. 6.6e-06;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 221 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTGCACCGT 266  
Db 1 GCTTGTACGCTGTACGCTGACGATAGATAGATAGATACACGTGCACCGT 46

## RESULT 4

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; CLONE: pTZgpt-F1s  
; IMMEDIATE SOURCE:  
US-08-232-463-14  
  
Query Match 13.3%; Score 41.6; DB 1; Length 7218;  
Best Local Similarity 4.2%; Pred. No. 0.0025;  
Matches 11; Conservative 152; Mismatches 101; Indels 0; Gaps 0;  
  
QY 15 GAAAGGGCCTCGGTAATACCATTTGTGGAAGAAAGTTGGTAATACGGAAGAAAGATTTCAT 74  
Db 1307 RRR 1248  
QY 75 CATCCAGAAAGGTGTGGAAGAAAGTTGTGGATTGCGTGGAAAGTTCGATCTGCACCATCT 134  
Db 1247 RRR 1188  
QY 135 CTAGATCGTGGAAAGTTACGTTAGCGCTTACGATACATATGTGATTGTGGAAGAAAGA 194  
Db 1187 RRR 1128  
QY 195 AGACGAGGACATCGGTGGAAGAAAGAAAGCTTGTACGCTGTACGCTGACGATAGATAGATAC 254  
Db 1127 RRR 1068  
QY 255 ACGTGACGCGCTCCACTTGACGCA 278  
Db 1067 RATCGCAGCTCCCTCGACCTGCA 1044

## RESULT 5

US-09-949-016-12417  
; Sequence 12417, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.

```
Query Match      11.7% ; Score 36.4; DB 3; Length 289;
Best Local Similarity 3.3%; Pred. No. 0.028;
Matches         7; Conservative          99; Mismatches 107; Indels    0; Gaps   0;
```

	Query Match	11.7%	Score 36.4;	DB 3;	Length 289;
	Best Local Similarity	3.3%;	Pred. No. 0.028;		
Matches	7; Conservative	99; Mismatches	107; Indels	0; Gaps	0
Qy	85	AGGTGCGAAAAAGTTGTGGATTGCCTGGAAAAGTTCGATCTGACCATCTCTAGATCGTG	144		
Dd	3	RGRGRACRABARUTURARCCEURARURURUPARCABARURURACRABARURGRRNRS	52		
Qy	145	GAAAAAGTTCCAGTTCAGCGCTTACGTACATATGTGGATTGGAAAAAAGAACGGCAGGC	204		
Dd	63	RNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRS	122		
Qy	205	ATCGGTGGAAGAAAGAAGCTTGTCAGCTCTACGCTCACGATAGATACACGTGCACGC	264		
Dd	123	RNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRS	182		
Qy	265	GTCCACTTGACGCAATTCACGCCAATGACG	297		
Dd	183	RNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNRNRSNR	215		

RESULT 8  
US-09-949-016-14020/c  
; Sequence 14020, Application US/09949016  
; Patent No. 6812339

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14020
; LENGTH: 22294
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14020
```

```
Query Match 10.8%; Score 33.8; DB 4; Length 22294;
Best Local Similarity 53.4%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 123 ATCTGACCACTCTTACATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGGGAT 182
DB 4766 AACTGTATTGTCAATGTCTTAAATGCGAAAAATTCATAGATTACATAACATGGAGAA 4707

QY 183 TGTGAAAAGAGACGGAGGACGTCGTGAAAAAGAACTGTACGCTGTACGCTGACG 242
DB 4706 TGAGGAAAATAAATAACAAGTAATGATGAAAAATGATGTGATCATTTAAAAAGCGGTG 4647

QY 243 ATAGATAGATACA 255
DB 4646 ATATAAATATAA 4634
```

```
RESULT 9
US-09-007-005-32
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
; US-09-007-005-32
```

```
Query Match 10.8%; Score 33.6; DB 3; Length 248;
Best Local Similarity 18.2%; Pred. No. 0.23;
Matches 28; Conservative 65; Mismatches 61; Indels 0; Gaps 0;

QY 136 TAGATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGGGATTTGGAAAAAGAA 195
DB 62 URGARARGARARCRARGARARCRURGRARURCRURGRARARCRURGRARARGRGRAR 121

QY 196 GACGGAGGATCGGTGGAAAAAGAACTGTACGCTGTACGCTGACGATAGATAGATACA 255
```

```
DB 122 CCRURGRURGRRCRGRURARARARCRURCRURGRARARCRARCRURGRARAR 181

QY 256 COTGCACGCGTCCACTTGACGCACCAATTGACGCA 289
DB 182 ARCRARCRARARCRURGRGRARARCRARGCR 215

RESULT 10
US-09-244-796-32
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
; US-09-244-796-32
```

```
Query Match 10.8%; Score 33.6; DB 3; Length 248;
Best Local Similarity 18.2%; Pred. No. 0.23;
Matches 28; Conservative 65; Mismatches 61; Indels 0; Gaps 0;

QY 136 TAGATCGTGGAAAAAGTTACGTTAGCGCTTACGTACATATGGGATTTGGAAAAAGAA 195
DB 62 URGARARGARARCRARGARARCRURGRARURCRURGRARARCRURGRARARGRGRAR 121

QY 196 GACGGAGGATCGGTGGAAAAAGAACTGTACGCTGTACGCTGACGATAGATAGATACA 255
DB 122 CCRURGRURGRRCRGRURARARARCRURCRURGRARARCRARCRURGRARAR 181

QY 256 COTGCACGCGTCCACTTGACGCACCAATTGACGCA 289
DB 182 ARCRARCRARARCRURGRGRARARCRARGCR 215
```

```
RESULT 11
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
```





TELEPHONE: 610-270-4478	
TELEFAX: 610-270-5090	
TELEX:	
INFORMATION FOR SEQ ID NO: 136:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1096 base pairs	
TYPE: nucleic acid	
STRANDEDNESS: single	
TOPOLOGY: linear	
US-08-858-207A-136	

QY	62	AAAGAAGATTCA	TCA	TCCAGAAAAGGTG	TGGAAAAGTTGT	TGGATTGCGT	GAAAAAGTTC	121
Db	6373	AGACAAGGTAAA	ATT	TACTTTGTTGAAA	AGATGTTT	TGAGAGTTTAT	CTGTGTTGAT	6314
QY	122	GATCTGACCATC	TCT	AGATCGTGGAAA	AGTTCA	CGTTAGCGCTT	ACGTACATAT	181
Db	6313	GTTTGACCAATC	TTT	TACTTTATGGAA	TTTAAA	CAGCGCAAGA	ATTAAAGTTT	6254
QV	182	TTGTGAAAAA	GAAGAC	CGGAGCATC	207			

Query Match	10.4%;	Score 32.4;	DB 3;	Length 1096;
Best Local Similarity	51.4%;	Pred. No. 1.2;		
Matches 75;	Conservative 0;	Mismatches 71;	Indels 0;	Gaps 0;
Qy	62	AAAGAAGATTTCATCCAGAAAAGGTGTGGAAAAGTTGTGGATTGCGTGGAAAAGTTC	121	
Db	699	AGACAAGGTAAATTTACTTTGAAAAGATGTTTGAGAGATTATCTGTGTGGTGAATCAAAATGA	640	
Qy	122	GATCTGACCATTCTAGATCGTGGAAAAGTTTCAGCTTGAGCCTTACCTACATATGTGGA	181	
Db	639	GTTTGAACAACTCTTTACTTTATGGAATTAACACAGCGGAGNAATTAAGTTCAGTTGTTTT	580	
Qy	182	TTGTGGAAAAAGAACGAGGCATC	207	
Db	579	TTCTGAAAAAATTTGCAGTTTCCGCATC	554	

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:55:42 ; Search time 1729.41 Seconds  
(without alignments)  
8741.717 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 gtcgaccatcattgaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	310.4	99.5	452	6	E64760	E64760 Chemically
2	310.4	99.5	452	6	AR427842	AR427842 Sequence
3	46	14.7	46	6	E64764	E64764 Chemically
4	46	14.7	46	6	E64765	E64765 Chemically
5	46	14.7	46	6	AR427845	AR427845 Sequence
6	46	14.7	46	6	AR427846	AR427846 Sequence
7	41.6	13.3	7218	6	I66494	I66494 Sequence 14
8	41	13.1	114519	8	AC124964	AC124964 Medicago
9	41	13.1	121565	2	AC135605	AC135605 Medicago
10	40.6	13.0	172623	2	AC147540	AC147540 Pan trogl
11	40.6	13.0	197004	2	AC145084	AC145084 Pan trogl
12	39.2	12.6	134442	8	AP004619	AP004619 Oryza sat
13	39.2	12.6	150300	8	AP003878	AP003878 Oryza sat
14	39.2	12.6	182932	2	AC112060	AC112060 Rattus no
15	39.2	12.6	185618	2	AC108612	AC108612 Rattus no
16	38.8	12.4	330362	2	AC129990	AC129990 Rattus no
17	36.8	11.8	63114	5	CR377227	CR377227 Zebraphish
18	36.8	11.8	139006	9	AL162431	AL162431 Human DNA
19	36.8	11.8	148068	9	CNS01RHN	AL161871 Human chr

RESULT 1  
E64760  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

ALIGNMENTS

E64760 452 bp DNA linear PAT 31-JAN-2002  
Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.  
E64760 GI:18628523  
JP 2000139477-A/1.  
synthetic construct  
other sequences; artificial sequences.  
1 (bases 1 to 452)  
Torii, R., Sawanto, S.B., Synge, P.K. and Gupta, S.K.  
Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
Patent: JP 2000139477-A 1 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH  
OS Artificial Sequence  
JP 2000139477-A/1  
PN 23-MAY-2000  
PD 23-MAY-2000  
PF 27-APR-1999 JP 1999119227  
PP 09-NOV-1998 IN 3322/98  
PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,  
PI SHIFU KUMAR GUPUTA  
PC C12N15/09, C12N5/10, C12N15/00, C12N5/00  
CC

Key Location/Qualifiers  
FT source 1.452  
FT /organism='Artificial Sequence'  
FEATURES  
source  
1.452  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'

ORIGIN  
Query Match 99.5%; Score 310.4; DB 6; Length 452;  
Best Local Similarity 99.7%; Pred. No. 6.4e-79;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCGACCATCATTTGAAAGGGCGCTCGGTAATACCATTTGGGAAAGTTGGTAATACGGA 60

AR162089 Sequence  
AR166614 Sequence  
AC131994 Mus muscu  
AC148087 Mus muscu  
Continuation (3 of  
AL031004 Arabidops  
AL132853 Caenorhab  
AL049607 Arabidops  
BX255947 Zebraphish  
AC021291 Homo sapi  
AC091549 Homo sapi  
AC090155 Homo sapi  
AL161579 Arabidops  
AC021184 Homo sapi  
Z95619 Caenorhabdi  
AC017912 Drosophil  
AB015945 Clostridi  
AC117968 Rattus no  
AC007054 Drosophil  
AC099213 Rattus no  
AC095337 Rattus no  
AC125836 Rattus no  
AE003785 Drosophil  
AC097988 Rattus no  
AC023297 Homo sapi  
BX537131 Danio rer

Db 1 GTGACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGA 60  
QY 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTTCGCTGGAAAAAGTT 120  
Db 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTTCGCTGGAAAAAGTT 120  
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCACGTTAGCGCTTACGTACATATGTGG 180  
Db 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCACGTTAGCGCTTACGTACATATGTGG 180  
QY 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACCGCTGA 240  
Db 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACCGCTGA 240  
QY 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCAACAATTGACGCAATGACGCCA 300  
Db 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCAACAATTGACGCAATGACGCCA 300  
QY 301 CTTGACGCTACT 312  
Db 301 CTTGACGCTACT 312

RESULT 2  
AR427842 LOCUS 452 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6639065.  
ACCESSION AR427842  
VERSION AR427842.1 GI:40186826  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 452)  
AUTHORS Tuli, R., Sawant, S.V., Singh, P.K. and Gupta, S.K.  
TITLE Chemically synthesized artificial promoter for high level expression of transgenes  
JOURNAL Patent: US 6639065-A 1 28-OCT-2003;  
FEATURES Location/Qualifiers  
source 1..452  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 99.5%; Score 310.4; DB 6; Length 452;  
Best Local Similarity 99.7%; Pred. No. 6.4e-79;  
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGA 60  
Db 1 GTGACCATCTTTGAAAGGCGCTCGGTATACCATTTGGAAAAAGTTGGTAAATACGA 60  
QY 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTTCGCTGGAAAAAGTT 120  
Db 61 AAAAGAAGATTTCATCATCCAGAAAAAGTGTGGAAAAAGTTGTGGATTTCGCTGGAAAAAGTT 120  
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCACGTTAGCGCTTACGTACATATGTGG 180  
Db 121 CGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCACGTTAGCGCTTACGTACATATGTGG 180  
QY 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACCGCTGA 240  
Db 181 ATTGTGAAAAAGAACGAGCGGCATCGTGGAAAAAGTGTGACGCTGTACCGCTGA 240  
QY 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCAACAATTGACGCAATGACGCCA 300  
Db 241 CGATAGATAGATACAGTGCACGCGTCCACTTGACGCAACAATTGACGCAATGACGCCA 300  
QY 301 CTTGACGCTACT 312  
Db 301 CTTGACGCTACT 312

RESULT 3  
E64764 LOCUS 46 bp DNA linear PAT 31-JAN-2002  
DEFINITION Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.  
ACCESSION E64764  
VERSION E64764.1 GI:18628527  
KEYWORDS JP 2000139477-A/5.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Tori, R., Sawanto, S.B., Synge, P.K. and Gupta, S.K.  
TITLE Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
JOURNAL Patent: JP 2000139477-A 5 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH  
COMMENT OS Artificial Sequence  
PN JP 2000139477-A/5  
PD 23-MAY-2000  
PF 27-APR-1999 JP 1999119227  
PI 09-NOV-1998 IN 3322/98  
PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,  
PI SHIFU KUMAR GUPUTA  
PC C12N15/09, C12N5/10, C12N15/00, C12N5/00  
CC  
FH Key Location/Qualifiers  
FT source 1..46  
/organism='Artificial Sequence'.  
FEATURES source  
1..46  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

ORIGIN  
Query Match 14.7%; Score 46; DB 6; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CCACCTTGACGCAACAATTGACGCAACAATTGACGCAACAATTGACGCTACT 312  
Db 1 CCACCTTGACGCAACAATTGACGCAACAATTGACGCAACAATTGACGCTACT 46

RESULT 4  
E64765 LOCUS 46 bp DNA linear PAT 31-JAN-2002  
DEFINITION Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same.  
ACCESSION E64765  
VERSION E64765.1 GI:18628528  
KEYWORDS JP 2000139477-A/6.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Tori, R., Sawanto, S.B., Synge, P.K. and Gupta, S.K.  
TITLE Chemically synthesized artificial promoter for realizing high-level expression of introduced gene and method for synthesizing the same  
JOURNAL Patent: JP 2000139477-A 6 23-MAY-2000;  
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH  
COMMENT OS Artificial Sequence  
PN JP 2000139477-A/6  
PD 23-MAY-2000  
PF 27-APR-1999 JP 1999119227  
PI 09-NOV-1998 IN 3322/98  
PI RAKESH TORI, SALLY BISHWANATO SAWANTO, PURAJUNNA KUMAR SYNGE,  
PI SHIFU KUMAR GUPUTA  
PC C12N15/09, C12N5/10, C12N15/00, C12N5/00

```
CC Key Location/Qualifiers
FH source 1..46
FT /organism='Artificial Sequence'.
FT Location/Qualifiers
FEATURES
  source 1..46
    /organism='synthetic construct'
    /mol_type='genomic DNA'
    /db_xref='taxon:32630'
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  221 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 266
  Db 1 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 46
RESULT 7
LOCUS I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
  source 1..7218
    /organism='unknown'
    /mol_type='unassigned DNA'
ORIGIN
  Query Match 13.3%; Score 41.6; DB 6; Length 7218;
  Best Local Similarity 4.2%; Pred. No. 0.52;
  Matches 11; Conservative 152; Mismatches 101; Indels 0; Gaps 0;
  15 GAAAGGCGCTCGGTATACCATTTGTGAAAAAGTTGGTAAACGAAAAAGAAAGATTTCAT 74
  Db 1307 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1248
  QY 75 CATCCAGAAAAGCTGTGGAAAAGTTGTGGATTGCGTGGAAAAGTTGCGATCTGACCATCT 134
  Db 1247 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1188
  QY 135 CTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACATATGTCGATTGTGAAAAAGA 194
  Db 1187 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1128
  QY 195 AGACGAGGACGATCGGTGGAAAAGAGCTTGTACGCTGACGTCGACGATAGATAGATAC 254
  Db 1127 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068
  QY 255 AGTGCACGCGTCCACTTGACGCA 278
  Db 1067 RATCGCAAGCTCCCTCGACCTGCA 1044
RESULT 8
LOCUS AC124964/c
DEFINITION Medicago truncatula clone mth2-27c4, complete sequence.
ACCESSION AC124964
VERSION AC124964.17 GI:45120181
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
REFERENCE 1 (bases 1 to 114519)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
TITLE Cook, D., Kim, D. and Roe, B.A.
JOURNAL Medicago truncatula BAC Clone mth2-27c4
REFERENCE 2 (bases 1 to 114519)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
  Cook, D., Kim, D. and Roe, B.A.
```

```
CC Key Location/Qualifiers
FH source 1..46
FT /organism='Artificial Sequence'.
FT Location/Qualifiers
FEATURES
  source 1..46
    /organism='synthetic construct'
    /mol_type='genomic DNA'
    /db_xref='taxon:32630'
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  221 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 266
  Db 1 GCTTGTACGCTGACGCTGACGATAGATAGATACACGTCGACGCGT 46
RESULT 5
LOCUS AR427845
DEFINITION Sequence 5 from patent US 6639065.
ACCESSION AR427845
VERSION AR427845.1 GI:40186829
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 46)
AUTHORS Tuli, R., Sawant, S.V., Singh, P.K. and Gupta, S.K.
TITLE Chemically synthesized artificial promoter for high level
  expression of transgenes
JOURNAL Patent: US 6639065-A 5 28-OCT-2003;
FEATURES
  source 1..46
    /organism='unknown'
    /mol_type='genomic DNA'
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
  Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  267 CCATTGACGCAATTTGACGCAATGACGCCACTTGACGCTACT 312
  Db 1 CCATTGACGCAATTTGACGCAATGACGCCACTTGACGCTACT 46
RESULT 6
LOCUS AR427846
DEFINITION Sequence 6 from patent US 6639065.
ACCESSION AR427846
VERSION AR427846.1 GI:40186830
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 46)
AUTHORS Tuli, R., Sawant, S.V., Singh, P.K. and Gupta, S.K.
TITLE Chemically synthesized artificial promoter for high level
  expression of transgenes
JOURNAL Patent: US 6639065-A 6 28-OCT-2003;
FEATURES
  source 1..46
    /organism='unknown'
    /mol_type='genomic DNA'
ORIGIN
  Query Match 14.7%; Score 46; DB 6; Length 46;
  Best Local Similarity 100.0%; Pred. No. 0.02;
```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 172623)  
Li, J. and Roe, B.A.  
Direct Submission  
Submitted (06-FEB-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Feb 6, 2004 this sequence version replaced gi:39573794.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2143: contig of 2143 bp in length  
2144 2243: gap of unknown length  
2244 4771: contig of 2528 bp in length  
4772 4872: gap of unknown length  
4872 7178: contig of 2307 bp in length  
7178 7278: gap of unknown length  
7278 9321: contig of 2043 bp in length  
9321 9422: gap of unknown length  
9422 12171: contig of 2750 bp in length  
12171 12272: gap of unknown length  
12272 14443: contig of 2172 bp in length  
14443 14544: gap of unknown length  
14544 17736: contig of 3193 bp in length  
17736 17837: gap of unknown length  
17837 21350: contig of 3514 bp in length  
21350 21450: gap of unknown length  
21450 24909: contig of 3459 bp in length  
24909 25009: gap of unknown length  
25009 28464: contig of 3455 bp in length  
28464 28564: gap of unknown length  
28564 32065: contig of 3501 bp in length  
32065 32185: gap of unknown length  
32185 36196: contig of 4031 bp in length  
36196 36297: gap of unknown length  
36297 41563: contig of 5267 bp in length  
41563 41664: gap of unknown length  
41664 45475: contig of 3812 bp in length  
45475 45576: gap of unknown length  
45576 48689: contig of 3114 bp in length  
48689 53726: contig of 4937 bp in length  
53726 53827: gap of unknown length  
53827 60447: contig of 6621 bp in length  
60447 60547: gap of unknown length  
60547 67085: contig of 6538 bp in length  
67085 67185: gap of unknown length  
67185 71628: contig of 4443 bp in length  
71628 71728: gap of unknown length  
71728 76314: contig of 4586 bp in length  
76314 76414: gap of unknown length  
76414 82810: contig of 6396 bp in length  
82810 82911: gap of unknown length  
82911 87615: contig of 4705 bp in length  
87615 87715: gap of unknown length  
87715 94295: contig of 6580 bp in length  
94295 94395: gap of unknown length  
94395 99025: contig of 4630 bp in length  
99025 99125: gap of unknown length  
99125 104801: contig of 5676 bp in length  
104801 104901: gap of unknown length  
104901 111798: contig of 6897 bp in length  
111798 111898: gap of unknown length  
111898 121365: contig of 9467 bp in length  
121365

\* 121366 121465: gap of unknown length  
\* 121466 129872: contig of 8407 bp in length  
\* 129872 129972: gap of unknown length  
\* 129972 140459: contig of 10487 bp in length  
\* 140459 140559: gap of unknown length  
\* 140559 140559: gap of unknown length  
\* 140560 149383: contig of 8824 bp in length  
\* 149383 149483: gap of unknown length  
\* 149483 159802: contig of 10319 bp in length  
\* 159802 159903: gap of unknown length  
\* 159903 172623: contig of 12721 bp in length.  
FEATURES  
Location/Qualifiers  
1. .172623  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="rp43-18c12"  
/clone\_lib="RPC1 - 43 Male Chimpanzee BAC Library"

## ORIGIN

Query Match 13.0%; Score 40.6; DB 2; Length 172623;  
Best Local Similarity 48.1%; Pred. No. 1.3;  
Matches 115; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 40 GGAAAAAGTTGGTAATACGGAAGAAAGAGATTTCATCCAGAAAGGTGTGAAAAGTT 99  
Db 107763 GGCAGACTTCCAGATACAAAAGGGTAGTGGGAATCTTGCCCATCTGGCATGGAGCGCT 107704  
QY 100 GTGATTGCGTGAAGAAAGTTCCATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGTT 159  
Db 107703 CAGCCTAGCATGCGCCCAAGCTGGAATCTTCTGCAAAAACAGGTCTGTAGCTTTTTCAT 107644  
QY 160 AGCGCTTACGTACATATGTGGATTGTGAAAAGAAAGAGCGAGGATCGGTGAAAAGA 219  
Db 107643 CATTTTAAAGTGAATTTGAGGAGTGTGGACACCCAGAGATCATTTTTCAGTCAATG 107584  
QY 220 AGCTTTAGCGTGTACGCTGACGATAGATAGATACACGTCGACGCGTCCACTTTGACGCA 278  
Db 107583 AATTTAAGCAAGATACGGAATAGTAGTATCCAGGTTCTCAGCTGCTATTGACCCA 107525

## RESULT 11

AC145064 197004 bp DNA linear HTG 23-APR-2004  
LOCUS Pan troglodytes clone rp43-45h9, WORKING DRAFT SEQUENCE, 11  
DEFINITION unordered pieces.  
ACCESSION AC145064  
VERSION AC145064.6 GI:46518614  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Pan troglodytes (Chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE 1 (bases 1 to 197004)  
AUTHORS Lau, C.C.Y. and Roe, B.A.  
TITLE Pan troglodytes BAC Clone rp43-45h9  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 197004)  
AUTHORS Lau, C.C.Y. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUN-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 197004)  
AUTHORS Lau, C.C.Y. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (23-APR-2004) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Apr 23, 2004 this sequence version replaced gi:40018761.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR







SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.		
REFERENCE	1		
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.	gene	/gene="OJ1120_C08.2"
TITLE	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC clone:OJ1120_C08	misc_feature	/note="probably inactive due to including frameshift(s) in CDS
JOURNAL	Published Only in Database (2001)		pseudogene, transposase"
REFERENCE	2 (bases 1 to 150300)		/pseudogene
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.		complement (14189, .19746)
TITLE	Direct Submission	gene	/gene="OJ1120_C08.3"
JOURNAL	Submitted (09-JUL-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan	misc_feature	complement(join(14189, .14750,14840, .14920,15600, .15671, 16144, .16221,17164, .17292,17710, .17790,17912, .18053, 18520, .18713,18810, .18875,18956, .19042,19174, .19240, 19556, .19746))
COMMENT	Tel:81-298-38-7441, Fax:81-298-38-7468)		/gene="OJ1120_C08.3"
	On Dec 2, 2003 this sequence version replaced gi:32335570.	CDS	/note="supported by full-length cDNA(s): AK111670"
	Genes were predicted from the integrated results of the following:		complement(join(14703, .14750,14840, .14920,15600, .15671, 16144, .16221,17164, .17292,17710, .17790,17912, .18053, 18520, .18713,18810, .18875,18956, .19042,19174, .19240))
	GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/cdb/glimmerm/glmr form.html), RiceHMM (http://www.tigr.org/software/glimmerm/), SLICEPredictor (http://rgp.dna.affrc.go.jp/RiceHMM/)_SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), Gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.	gene	/note="contains EST(s): C97094(C51661), AU108860(C51661) contains full-length cDNA(s): AK111670"
	A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.	misc_feature	/codon_start=1
	The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OJ1120_C08 clone has an overlap with P0538806 (DDBJ: AP004619) clone at 5' end and with P0577G06 (DDBJ: AP004590) clone at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.	gene	/product="putative RAE1 (RNA export 1, S.pombe) homolog"
FEATURES	Location/Qualifiers		/protein_id="BAD03038.1"
source	1. .150300		/db_xref="GI:3863679"
	/organism="Oryza sativa (japonica cultivar-group)"		/translation="MATGLAPNMNPNKSFVLPNPGDSVSLSPKSNLLVATSWD NOVRCWEIVGGSSOPKASISHDQPVLCASAKWDDGTVFSGCDKVKWPLLSGGQAO TVAMHDAVKSIAMIPONNLLVSGSWDKTLRYWDTROSNPVHVQQLPERCYALTYNYP LMIYGTARNLVFNQLQNPQTEFKRIQSLFKITQKCLAAFPDQQGLVGSIEGRVGVH LKVGTSKQNFTHKQREGNDFSVNSLNFHPVHHFTAFAGSDGAFNFWDKDKQRLK AFSCPLPFGSTFNSDGSIFAYAVCYDWSRGAENHPATATSIYLSHSPQSEVYKKG PRATSRK"
	/mol_type="genomic DNA"		complement (21178, .21432)
	/cultivar="Nipponbare"		/gene="OJ1120_C08.4"
	/db_xref="taxon:39947"		complement (21178, .21432)
	/chromosomes="8"		/gene="OJ1120_C08.4"
	/clone="OJ1120_C08"		/note="hypothetical ORF predicted by GlimmerM
gene	5270, .8150		this category is not included in IRGSP standard"
misc_feature	5270, .8150		join(21808, .21835,21998, .22197)
	/genes="OJ1120_C08.1"		/gene="OJ1120_C08.5"
	/genes="OJ1120_C08.1"		join(21808, .21835,21998, .22197)
	/note="probably inactive due to 3' exon missing in CDS		/gene="OJ1120_C08.5"
	probably inactive due to 5' exon missing in CDS		/note="hypothetical ORF predicted by GlimmerM
pseudogene	MLA6 protein"		this category is not included in IRGSP standard"
pseudogene	10283, .12816		complement(join(23249, .24465,24547, .24604,24703, .24780, 24880, .24939,25067, .25180,25284, .25730))
gene	/genes="OJ1120_C08.2"		/gene="OJ1120_C08.7"
misc_feature	10283, .12816		complement (join(23249, .24465,24547, .24604,24703, .24780, 24880, .24939,25067, .25180,25284, .25730))
			/gene="OJ1120_C08.7"
			complement (join(23249, .24465,24547, .24604,24703, .24780, 24880, .24939,25067, .25180,25284, .25730))
			/gene="OJ1120_C08.7"
			/note="supported by full-length cDNA(s): AK107028"
			23501, .24827
			<23501, .>24827
			/gene="OJ1120_C08.6"
			/note="supported by full-length cDNA(s): AK107399"
			23501, .24827
			/gene="OJ1120_C08.6"
			/note="supported by full-length cDNA(s): AK107399"
			23501, .24827
			/gene="OJ1120_C08.6"
			/note="contains full-length cDNA(s): AK107399
			non-coding transcript
			probably inactive due to including stop codon(s) in CDS"
			complement(join(23819, .24465,24547, .24604,24703, .24780, 24880, .24939,25067, .25180,25284, .25496))
			/gene="OJ1120_C08.7"
			/note="contains full-length cDNA(s): AK107028, AK109294"
			/codon_start=1
			/product="putative transcription factor STGAL"
			/protein_id="BAD03039.1"
			/db_xref="GI:38636798"
			/translation="MTSAPAAQAFAPAPAMGIYDRRHHPPLAAGVGDHPFIRPDT TASTNSAAAAAMVAVPPLTPKFSQALPLQHGDDQDAAALQESPRHADSFEQE ASKPRDKIQRLRAQNREARKSLRKAYIQNLTSRMKLALHLEQETIRARQOSYAIN RSSNPATLPAPIDSGVTVTFEYEAQWVEEQGQTAEELRASLQAAAEGLRAVVEAAL

```

AHYDRFLFAAKREARDVFFVMSGVWRTGAERFFLWIAGFRPSEVIRVIAPOLEPMT
ROADYVGGQKARHEDALSGQMDLKTQTLADSLAEAVVWSTSDCHAPPPPPPEE
EPSSAAGDGGCYMAQMGSLNLFVAFVHRHRRSPPTSHLHVRRAELG"
gene      complement (join(29680..29827,31003..31098,32215..32219,
33217..33312))
/misc_feature  /genes="OJ1120 C08.8"
complement (join(29680..29827,31003..31098,32215..32219,
33217..33312))
/notes="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
gene      join(37583..37663,37864..38022)
/misc_feature  /genes="OJ1120 C08.9"
join(37583..37663,37864..38022)
/notes="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
gene      complement (38722..38910)
/genes="OJ1120 C08.10"
mRNA      complement (<38722..>38910)
/genes="OJ1120 C08.10"
/notes="start and end point are not identified"
complement (38722..38910)
/genes="OJ1120 C08.10"
/notes="predicted by FGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BA003040.1"
/db_xref="GI:38636799"
/translations="MYVEAAVSSLKQQQQQQQGHGSAARSHQVSIKEKQQQQIG
CRTLAPRIYPTASGKK"
gene      complement (38954..39154)
/misc_feature  /genes="OJ1120 C08.11"
complement (38954..39154)
/notes="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
gene      complement (join(41298..41610,42041..43471,44688..45078,
45161..45373,45480..45520))
/genes="OJ1120 C08.12"
mRNA      complement (join(41298..41610,42041..43471,44688..45078,
45161..45373,45480..45520))
/genes="OJ1120 C08.12"
/notes="supported by full-length cDNA(s): AK105561"
complement (join(42104..43471,44688..45078,45161..45351))

Query Match      12.6%; Score 39.2; DB 8; Length 150300;
Best Local Similarity 50.0%; Pred. No. 3.2;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 29 AATACCATTTGGAAAAAGTTGGTATACGGAAGAAAGAGATTCATCCAGAAAGGT 88
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83780 AAAAAAAGTAAATAAAGATTGCTATTGCTAAGGATTGCAAGATAGTACAAAAAAGG 83721
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 89 GTGGAAGTTGTTGGTTGGTGGAAAGATTTCGATTCGACATCTCTAGATCGTGGAA 148
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83720 GTAACAAGTTGTGTCGGAGTGGAAACAAATTTGTTGAGTGGACCTTATTTGGAATTA 83661
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 149 AAGTTCACGTTAGCGCTTACATATGTCGATTGTGGAAGAAAGACGGAGGCATCG 208
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83660 AAGAGAAATTACAGGGGCAATTTTGTTCATTGTGAAAAAGTAATAAATGAAG 83601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 209 GTGGAAGAAAGACTT 224
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83600 GATTGGAAGAAAGCTT 83585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AC112060/c
LOCUS      AC112060      182932 bp      DNA      linear      HTG 15-NOV-2002
```

DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Rattus norvegicus clone CH230-180E21, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 3 unordered pieces.  
AC112060  
GI:25006759  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 182932)  
Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Aisbrooks, S., Amin, A., Arguliano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaraike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, R.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevarra, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jacks, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensushewa, L., Loulseged, H., Lozano, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczky, R., Wooden, H., Wortley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 182932)  
Worley, K. C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 182932)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:22856848.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: QGFs  
Center clone name: CH230-180E21

----- Summary Statistics

Assembly program: phrap; version 0.990329  
Consensus quality: 171712 bases at least Q40  
Consensus quality: 174959 bases at least Q30  
Consensus quality: 177170 bases at least Q20  
Estimated insert size: 183207; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 179536: contig of 179536 bp in length  
\* 179537 179636: gap of unknown length  
\* 179637 181302: contig of 1666 bp in length  
\* 181303 181402: gap of unknown length  
\* 181403 182932: contig of 1530 bp in length.

## FEATURES

source  
1. 182932  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-180E21"  
misc\_feature  
1. 1190  
/notes="wgs\_contig"  
misc\_feature  
177608..179536  
/note="wgs\_contig"

## ORIGIN

Query Match 12.6%; Score 39.2; DB 2; Length 182932;  
Best Local Similarity 49.5%; Pred. No. 3.2;  
Matches 101; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
Qy 17 AAGGGCTCGTATACCATTTGGAAAAAGTTGGTAATACGAAAAAGAGATTTCATCA 76  
Db 180078 AAAGACATGAAGAACTCCCTTAGGAAACACAGGAAACATTAATAACAGTAGAAGCC 180019  
Qy 77 TCCAGAAAAGGTGTGAAAAGTTGTGGATTGGCGTGGAAAAAGTTTCGATCTGACCTCTCT 136  
Db 180018 TACAGAGAGAAATCGCAAAATCCCTGAAGAAATCGCAAAATGCTGAAAGAAATCCA 179959  
Qy 137 AGATCGTGGAAAAAGTTACGTTAGCGCTTTACGTACATATGTGGATTGTGGAAAAAGAG 196  
Db 179958 GGAACACACAAACAGTTGAAGGAACATAAAATGGAATAGAACATCAAGAAAGAACAC 179899

Qy 197 ACAGGAGCATCGTGGAAAAAGAA 220  
Db 179898 ATGGAACAACACCTCGATATAGAA 179875  
RESULT 15  
AC108612  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-293B15, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC108612  
AC108612.5 GI:25139237  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus.  
1. (bases 1 to 185618)  
Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Greggorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunsarane, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuhewa, L., Louisedge, H., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,  
Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Direct Submission  
Unpublished  
2 (bases 1 to 185618)  
Worley, K. C.  
Direct Submission  
TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 185618)  
Worley, K. C.  
AUTHORS  
TITLE



Page blank (usp:0)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:53:27 ; Search time 474.353 Seconds  
(without alignments)

3893.639 Million cell updates/sec

Title: US-10-814-858A-1

Perfect score: 312

Sequence: 1 gtcgaccatttgaaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002as:\*  
7: Geneseq2002bs:\*  
8: Geneseq2003as:\*  
9: Geneseq2003bs:\*  
10: Geneseq2003cs:\*  
11: Geneseq2003ds:\*  
12: Geneseq2004as:\*  
13: Geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	310.4	99.5	452	3	AAA28449
2	46	14.7	46	3	AAA28438
3	46	14.7	46	3	AAA28437
4	36.8	11.8	110000	11	ACW45090_1
5	34.4	11.0	7450	4	ABL02730
6	33	10.6	4074	10	ADB69083
7	32.6	10.4	468	6	ABK76259
8	32.6	10.4	110000	6	ABQ69245_01
9	32.6	10.4	110000	6	ABQ67197_00
10	32.4	10.4	495	10	ABX05935
11	32.4	10.4	1096	2	AAT98670
12	32.4	10.4	1727	2	AAZ96371
13	32.4	10.4	2992	4	AAI66078
14	32.4	10.4	6380	6	ABN80109
15	32.4	10.4	11303	2	AAV52248
16	32.4	10.4	65454	12	ADN01773
17	32.4	10.4	83517	13	ABD32581
18	32.4	10.4	110000	10	ABS56454_01
19	32	10.3	110000	6	ABA03041_22
20	31.8	10.2	722	6	ABQ68303

#### ALIGNMENTS

##### RESULT 1

AAA28449

ID AAA28449 standard; DNA; 452 BP.

XX AAA28449;

XX

DT 29-AUG-2000 (first entry)

DE Artificial synthetic promoter for high level plant gene expression.

KW Artificial promoter; strong; weak; transgene expression; plant; ss.

XX

XX Synthetic.

XX

XX Key

FT misc\_signal

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Abq70825 Listeria  
Aca28221 Prokaryot  
Aaa42022 Human sec  
Abx56693 Arabidops  
Abl09683 Drosophil  
Abl19785 Drosophil  
Aav09296 Nucleotid  
Aad05761 Arabidops  
Aad05787 Arabidops  
Aad05648 A. thalia  
Abk5208 Arabidops  
Add55841 Thalecres  
Adi41682 Plant tra  
Ado2050 Thalecres  
Ado03392 Thalecres  
Abl09682 Drosophil  
Abl19784 Drosophil  
Aab40014 Human che  
Aas20588 Human met  
Abd33020 Mouse can  
Aah00080 Enterococ  
Ada71938 Rice gene  
Abl34132 Human imm  
Aba21231 Human ner

#### ALIGNMENTS

##### RESULT 1

AAA28449

ID AAA28449 standard; DNA; 452 BP.

XX AAA28449;

XX

DT 29-AUG-2000 (first entry)

DE Artificial synthetic promoter for high level plant gene expression.

KW Artificial promoter; strong; weak; transgene expression; plant; ss.

XX

XX Synthetic.

XX

XX Key

FT misc\_signal

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

```
PF 25-FEB-1999; 99EP-00301419.
XX
PR 09-NOV-1998; 98IN-DE003322.
XX
PA (COUL ) CSIR COUNCIL SCI IND RES.
PI Tuli R, Sawant SV, Singh PK, Gupta SK;
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
PT
XX Claim 13; Page 11; 40pp; English.
XX
XX Chemically synthesized artificial promoters are new and comprise a DNA
CC sequence designed for a targeted level and pattern of gene expression by
CC strategically putting together several signature sequences identified by
CC sequence alignment and statistical analysis of a large database
CC constructed for this purpose. A method for chemically synthesizing an
CC artificial promoter for expressing genes at a desired level in different
CC organisms is also claimed. The high level expression in a plant using
CC such an artificial promoter (e.g. AAA28449) can be measured comprising
CC polyethylene glycol (PEG) mediated transformation of plant protoplasts as
CC well as biolistic mediated transformation of plant tissues including
CC root, stem, intact leaf tissue followed by transient GUS assay to compare
CC with a natural CamV 35S promoter showing the desired level of activity.
CC The promoter is useful for high level expression of transgenes in
CC different organisms and for testing high level gene expression in plants
CC (claimed). The promoter is biologically active and is efficient and can
CC be synthesized to express in even the most complex organisms
XX
XX Sequence 452 BP; 157 A; 88 C; 95 G; 112 T; 0 U; 0 Other;
SQ
Query Match 99.5%; Score 310.4; DB 3; Length 452;
Best Local Similarity 99.7%; Pred. No. 2e-87;
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCGACCATCATTTGAAAGGCCCTCGGTATATACATTTGTGGAAGTTGGTAATACGGA 60
DB 1 GTCGACCATCATTTGAAAGGCCCTCGGTATATACATTTGTGGAAGTTGGTAATACGGA 60
QY 61 AAAAGAGATTTCATCATCAGAAAGGTGTGGAAGTTCTGATTCGCTGGAAAAGTT 120
DB 61 AAAAGAGATTTCATCATCAGAAAGGTGTGGAAGTTCTGATTCGCTGGAAAAGTT 120
QY 121 CGATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACATATGTGG 180
DB 121 CGATCTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACATATGTGG 180
QY 181 ATTGTGGAAGAAAGACGAGCGCATCGGTGGAAAAGCTTGTACGCTGTACGCTGA 240
DB 181 ATTGTGGAAGAAAGACGAGCGCATCGGTGGAAAAGCTTGTACGCTGTACGCTGA 240
QY 241 CGATAGATAGATACGTCACGCGTCCACTTCACCCACATTTGACGCAATGACGCCA 300
DB 241 CGATAGATAGATACGTCACGCGTCCACTTCACCCACATTTGACGCAATGACGCCA 300
QY 301 CTTGACGCTACT 312
DB 301 CTTGACGCTACT 312
```

## RESULT 2

AAA28438

ID AAA28438 standard; DNA; 46 BP.

AC AAA28438;

XX 29-AUG-2000 (first entry)

XX Synthetic promoter conserved domain I.

XX

```
KW Artificial promoter; strong; weak; transgene expression; plant; ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT misc_signal 1..21
FT /tag= a
FT /label= U_box
FT /note= "GTACGC type element"
FT repeat_region 22..33
FT /tag= b
FT /rpt_type= TANDEM
FT repeat_unit 22..25
FT /tag= c
FT /note= "GATA-type cis-acting element"
FT GC_signal 34..46
FT /tag= d
FT /note= "Resembles GC-box, may play role in kinetics of
FT opening of the transcription bubble and keeping the
FT minimal promoter in a most active form to enhance
FT transcription reinitiation"
XX
XX EP1002869-A1.
XX
XX 24-MAY-2000.
XX
XX 25-FEB-1999; 99EP-00301419.
XX
XX 09-NOV-1998; 98IN-DE003322.
XX
XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
XX Tuli R, Sawant SV, Singh PK, Gupta SK;
XX
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
PT expression of transgenes in different organisms.
XX
XX Claim 5; Page 16; 40pp; English.
XX
XX A chemically synthesized promoter can comprise a conserved domain I as
CC shown here for high level expression of genes. Chemically synthesized
CC artificial promoters are new and comprise a DNA sequence designed for a
CC targeted level and pattern of gene expression by strategically putting
CC together several signature sequences identified by sequence alignment and
CC statistical analysis of a large database constructed for this purpose. A
CC method for chemically synthesizing an artificial promoter for expressing
CC genes at a desired level in different organisms is also claimed. The high
CC level expression in a plant using such an artificial promoter (e.g.
CC AAA28449) can be measured comprising polyethylene glycol (PEG) mediated
CC transformation of plant protoplasts as well as biolistic mediated
CC transformation of plant tissues including root, stem, intact leaf tissue
CC followed by transient GUS assay to compare with a natural CamV 35S
CC promoter showing the desired level of activity. The promoter is useful
CC for high level expression of transgenes in different organisms and for
CC testing high level gene expression in plants (claimed). The promoter is
CC biologically active and is efficient and can be synthesized to express in
CC even the most complex organisms
XX
XX Sequence 46 BP; 11 A; 11 C; 13 G; 11 T; 0 U; 0 Other;
SQ
Query Match 14.7%; Score 46; DB 3; Length 46;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 221 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTCACGCGT 266
DB 1 GCTTGTACGCTGTACGCTGACGATAGATAGATACACGTCACGCGT 46
```

RESULT 3

AAA28437



AA28437 standard; DNA; 46 BP.  
AC AAA28437;  
DT 29-AUG-2000 (first entry)  
DE Synthetic promoter minimal domain (b) for high level gene expression.  
KW Artificial promoter; strong; weak; transgene expression; plant; ss.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT CAAT\_signal 1..5  
FT /\*tag= a  
FT /note= "conserved"  
FT misc\_signal 6..10  
FT /\*tag= b  
FT /note= "conserved TGACG box"  
FT CAAT\_signal 11..16  
FT /\*tag= c  
FT /note= "conserved"  
FT misc\_signal 17..21  
FT /\*tag= d  
FT /note= "conserved TGACG box"  
FT CAAT\_signal 22..27  
FT /\*tag= e  
FT /note= "conserved"  
FT misc\_signal 27..31  
FT /\*tag= f  
FT /note= "conserved TGACG box"  
FT CAAT\_signal 32..36  
FT /\*tag= g  
FT /note= "conserved"  
FT misc\_signal 37..41  
FT /\*tag= h  
FT /note= "conserved TGACG box"  
FT CAAT\_signal 42..46  
FT /\*tag= i  
FT /note= "conserved"  
EP1002869-A1.  
XX  
PN  
XX  
XX  
PD 24-MAY-2000.  
XX  
XX  
PF 25-FEB-1999; 99EP-00301419.  
XX  
XX  
PR 09-NOV-1998; 98IN-DE003322.  
XX  
XX  
PA (COUL ) CSIR COUNCIL SCI IND RES.  
XX  
XX  
PI Tuli R, Sawant SV, Singh PK, Gupta SK;  
XX  
XX  
DR WPI; 2000-341712/30.  
XX  
XX  
PT New chemically synthesized artificial promoter, useful high level  
PT expression of transgenes in different organisms.  
XX  
PS Claim 4; Page 15; 40pp; English.  
XX  
XX  
CC A chemically synthesized promoter can comprise a minimal domain (b) as  
CC shown here for high level expression of genes. This domain has a number  
CC of conserved sequences interspersed with a TGACG box. Chemically  
CC synthesized artificial promoters are new and comprise a DNA sequence  
CC designed for a targeted level and pattern of gene expression by  
CC strategically putting together several signature sequences identified by  
CC sequence alignment and statistical analysis of a large database  
CC constructed for this purpose. A method for chemically synthesizing an  
CC artificial promoter for expressing genes at a desired level in different  
CC organisms is also claimed. The high level expression in a plant using  
CC such an artificial promoter (e.g. AA28449) can be measured comprising  
CC polyethylene glycol (PEG) mediated transformation of plant protoplasts as  
CC well as biolistic mediated transformation of plant tissues including

CC root, stem, intact leaf tissue followed by transient GUS assay to compare  
CC with a natural CaMV 35S promoter showing the desired level of activity.  
CC The promoter is useful for high level expression of transgenes in  
CC different organisms and for testing high level gene expression in plants  
CC (claimed). The promoter is biologically active and is efficient and can  
CC be synthesized to express in even the most complex organisms  
XX  
SQ Sequence 46 BP; 13 A; 16 C; 8 G; 9 T; 0 U; 0 Other;  
Query Match 14.7%; Score 46; DB 3; Length 46;  
Best Local Similarity 100.0%; Pred. No. 0.00017;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 267 CCACCTTGACGCAACAATTGACGCAATGACGCCACTTGACGCTACT 312  
DB 1 CCACTTGACGCAACAATTGACGCAATGACGCCACTTGACGCTACT 46  
RESULT 4  
ACN45090\_1  
Continuation (2 of 4) of ACN45090 from base 100001 (Human genomic sequence HCG22125.)  
WP Sequence split into 4 fragments LOCUS ACN45090 Accession ACN45090  
WP Fragment Name Begin End  
WP ACN45090\_0 1 110000  
WP ACN45090\_1 100001 210000  
WP ACN45090\_2 200001 310000  
WP ACN45090\_3 300001 350764  
Query Match 11.8%; Score 36.8; DB 11; Length 110000;  
Best Local Similarity 56.7%; Pred. No. 2.7;  
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 70 TTCATCATCCAGAAAGGTGTGAAAAGTTGTGGATTGCGTGGAAAAGTTTCGATCTGAC 129  
DB 76489 TTCTTTTAAAATAATGATGCAAAAGCTGTGTTGGCGGGGTATTTTACATCAGAA 76548  
QY 130 CATCTCTAGATCGTGGAAAAGTTACGTTAGCGCTTACGTACATATGTGGATTGTGGAA 189  
DB 76549 ATGCTTTAAGAGATTGTAATAAGGTTGTTGTAAGAACTGAGGTACATTTAATGAATATGAA 76608  
RESULT 5  
ABL02730  
ID ABL02730 standard; cDNA; 7450 BP.  
XX  
XX  
AC ABL02730;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2672.  
XX  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
XX  
PN WO200171042-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB58627.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more



CC in expression of genes may be used to provide a representation of the way  
CC in which *Bacillus* cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 468 BP; 124 A; 112 C; 114 G; 118 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.6; DB 6; Length 468;

Best Local Similarity 55.9%; Pred. No. 7;

Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 21 GCCTCGGTATACATTCGGAAGAGTTGGTATACGGAAGAGAGATTCATCATCCA 80

Db 191 GTCTCGAACAGTCGGCTTTACACGATTCAAAATGCGGAAGAGCGATTTTATATCC 250

QY 81 GAAAGAGTGTGGAAGAGTTGCGATTGCGTGGAAAAAGTTTCGATCTGACCA 131

Db 251 GGAAGGGCTTGAACTTTAGACAGTCCATGGAAATATGGGAGCTGACAA 301

#### RESULT 8

ABQ69245\_01  
Continuation (2 of 31) of ABQ69245 from base 100001 (*Listeria innocua* DNA sequence #684).

WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245

WP Fragment Name Begin End

WP ABQ69245\_00 1 110000

WP ABQ69245\_01 100001 210000

WP ABQ69245\_02 200001 310000

WP ABQ69245\_03 300001 410000

WP ABQ69245\_04 400001 510000

WP ABQ69245\_05 500001 610000

WP ABQ69245\_06 600001 710000

WP ABQ69245\_07 700001 810000

WP ABQ69245\_08 800001 910000

WP ABQ69245\_09 900001 1010000

WP ABQ69245\_10 1000001 1110000

WP ABQ69245\_11 11000001 1210000

WP ABQ69245\_12 12000001 1310000

WP ABQ69245\_13 13000001 1410000

WP ABQ69245\_14 14000001 1510000

WP ABQ69245\_15 15000001 1610000

WP ABQ69245\_16 16000001 1710000

WP ABQ69245\_17 17000001 1810000

WP ABQ69245\_18 18000001 1910000

WP ABQ69245\_19 19000001 2010000

WP ABQ69245\_20 20000001 2110000

WP ABQ69245\_21 21000001 2210000

WP ABQ69245\_22 22000001 2310000

WP ABQ69245\_23 23000001 2410000

WP ABQ69245\_24 24000001 2510000

WP ABQ69245\_25 25000001 2610000

WP ABQ69245\_26 26000001 2710000

WP ABQ69245\_27 27000001 2810000

WP ABQ69245\_28 28000001 2910000

WP ABQ69245\_29 29000001 3010000

WP ABQ69245\_30 30000001 3111208

#### Query Match

Best Local Similarity 10.4%; Score 32.6; DB 6; Length 110000;

Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 62 AAGAAGATTCATCCAGAAAAGTGTGGAAGAGTTGTGGATTGCGTGGAAAAAGTTTC 121

Db 46002 AATGAGGTTTAAATCAAAATCGGGTGAAGAGATGAATGAAGATGAAGAGCTT 46061

QY 122 GATCTGACCATCTCTAGATCGTGGAAAAAGTTTACGCTTACGTTACATATGTGA 181

Db 46062 TTTAGTATTACTTCTGTTGTTGTAGTATTATTCGTTAGTGTGTTGATCAATTGTGCA 46121

QY 182 TTGTGGAAGAAAGACGAGGCGG 204

Db 46122 GAGTCAAAACGGGCGGAAGCGGC 46144

#### RESULT 9

ABQ67197\_00

WP Sequence split into 12 fragments LOCUS ABQ67197 Accession Abq67197

WP Fragment Name Begin End

WP ABQ67197\_00 1 110000

WP ABQ67197\_01 100001 210000

WP ABQ67197\_02 200001 310000

WP ABQ67197\_03 300001 410000

WP ABQ67197\_04 400001 510000

WP ABQ67197\_05 500001 610000

WP ABQ67197\_06 600001 710000

WP ABQ67197\_07 700001 810000

WP ABQ67197\_08 800001 910000

WP ABQ67197\_09 900001 1010000

WP ABQ67197\_10 1000001 1110000

WP ABQ67197\_11 1100001 1163020

ID ABQ67197 standard; DNA; 1163020 BP.

XX AC

XX ABQ67197;

XX 29-AUG-2002 (first entry)

XX *Listeria innocua* contig DNA sequence #10.

DE *Listeria innocua* contig DNA sequence #10.

XX Antibacterial; *Listeria*; food contamination; mutational analysis;

KW Infection; ds.

XX *Listeria innocua*.

OS *Listeria innocua*.

XX WO200228891-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-FR003061.

XX 04-OCT-2000; 2000FR-00012697.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX New genomic sequences from *Listeria* species, useful for detection,

XX treatment and prevention of infection, also related polypeptides,

XX antibodies and modulators.

XX Claim 5; SEQ ID NO 10; 180bp; French.

XX The present invention relates to nucleic acid sequences (ABQ67189-

XX ABQ71212) from *Listeria* sp. The sequences are useful as probes and

XX primers for identification and/or detection of *Listeria* (e.g. as

XX contaminants in foods, or mutational analysis) and for analysis of gene

XX expression. Proteins encoded by the nucleic acid sequences can be used to

XX screen for compounds that modulate gene expression, replication and

XX pathogenicity of *Listeria* (potential therapeutic agents), also for

XX treating infections by *Listeria*, and are useful as immunogens in anti-

XX *Listeria* vaccines. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1163020 BP; 388339A; 197589C; 235612G; 341474T; 0U; 60Other;

Query Match 10.4%; Score 32.6; DB 6; Length 110000;

Best Local Similarity 51.7%; Pred. No. 57;

Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;



XX This sequence encodes a Streptococcus pneumoniae protein that (based on  
 CC homology with a Bacillus subtilis protein) is a BSELABCD NCBI gi:  
 CC 895746, and represents a DNA sequence of the invention. The DNA sequences  
 CC were isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794).  
 CC The Streptococcus pneumoniae proteins of the invention can be used to  
 CC identify compounds which interact with and inhibit or activate the  
 CC activity of the proteins. Antagonists can be used to treat diseases  
 CC caused by S. pneumoniae proteins, through genetic immunisation. They can  
 CC also be used to induce an immunological response in a mammal by  
 CC inoculation with the S. pneumoniae proteins or delivery of the encoding  
 CC nucleic acids in a vector adequate to produce antibody and/or T cell  
 CC immune responses to protect the animal from disease. The proteins can  
 CC also be used to identify antimicrobial compounds which are capable of  
 CC inhibiting their bioactivity. In particular the proteins of the invention  
 CC can be used to prevent adhesion of bacteria to mammalian extracellular  
 CC matrix proteins on in-dwelling devices or in wounds, to block protein-  
 CC mediated mammalian cell invasion, and to block the normal progression of  
 CC pathogenesis in infections initiated other than by the implantation of in  
 CC -dwelling devices or other surgical techniques  
 CC  
 XX Sequence 1096 BP; 338 A; 224 C; 138 G; 396 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.4; DB 2; Length 1096;  
 Best Local Similarity 51.4%; Pred. No. 11;  
 Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
 QY 62 AAGAAGATTCATCCAGAAAGGTGTGGAAGTTTGGATTGGTGGAAAGATTC 121  
 DB 699 AGACAAGGTAAATTTACTTGAAGAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 640  
 QY 122 GATCTGACCATCTCTAGATCGTGGAAAAGTTTCAACGTTAGCGTTACGTACATATGTGGA 181  
 DB 639 GTTTGAACAATCTTTACTTTATGGAATTAACAGCGGAGAGTAAGTTCAGTTGTTTT 580  
 QY 182 TTGTGGAAGAAAGACGGAGGCATC 207  
 DB 579 TTCTGAAAAAATTCGATCCGCATC 554

RESULT 12  
 AAZ96371/c  
 ID AAZ96371 standard; DNA; 1727 BP.  
 XX  
 AC AAZ96371;  
 XX  
 DT 10-APR-2000 (first entry)  
 XX  
 DE S. pneumoniae derived DNA from ORF #199.  
 XX  
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;  
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9806734-A1.  
 XX  
 PD 19-FEB-1998.  
 XX  
 PF 15-AUG-1997; 97WO-US014436.  
 XX  
 PR 16-AUG-1996; 96US-0024022P.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 PI Stodola RK;  
 XX  
 DR WPI; 1998-159452/14.  
 DR P-PSDB; AAY86040.  
 XX  
 PT Streptococcus pneumoniae proteins and related DNA - useful for screening  
 PT compounds for antibacterial activity.

XX Claim 4; Page 230-231; 640pp; English.  
 XX  
 CC This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see  
 CC AA85792-Y86182). The DNA, vectors and host cells described in the method  
 CC of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful for  
 CC inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease  
 XX  
 SQ Sequence 1727 BP; 546 A; 332 C; 221 G; 628 T; 0 U; 0 Other;

Query Match 10.4%; Score 32.4; DB 2; Length 1727;  
 Best Local Similarity 51.4%; Pred. No. 13;  
 Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
 QY 62 AAGAAGATTCATCCAGAAAGGTGTGGAAGTTTGGATTGGTGGAAAGATTC 121  
 DB 696 AGACAAGGTAAATTTACTTGAAGAGATGGTTGAGAGTTTATCTGTGGTGAATCAATGA 637  
 QY 122 GATCTGACCATCTCTAGATCGTGGAAAAGTTTCAACGTTAGCGTTACGTACATATGTGGA 181  
 DB 636 GTTGAACAATCTTTACTTTATGGAATTAACAGCGGAGAGTAAGTTCAGTTGTTTT 577  
 QY 182 TTGTGGAAGAAAGACGGAGGCATC 207  
 DB 576 TTCTGAAAAAATTCGATCCGCATC 551

RESULT 13  
 AA166078  
 ID AA166078 standard; DNA; 2992 BP.  
 XX  
 AC AA166078;  
 XX  
 DT 14-JAN-2002 (first entry)  
 XX  
 DE Spinacia oleracea ClpC protease gene (GenBank: AF043539).  
 XX  
 KW Transgenic plant; stress tolerance; heat shock protein; HSP; cotton;  
 KW canola; soybean; corn; wheat; tobacco; sorghum; potato; tomato;  
 KW Arabidopsis thaliana; ds.  
 XX  
 OS Spinacia oleracea.  
 XX  
 PN WO200170929-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 20-MAR-2001; 2001WO-US008836.  
 XX  
 PR 20-MAR-2000; 2000US-0190769P.  
 PR 18-APR-2000; 2000US-0198116P.  
 XX  
 PA (ARCH-) ARCH DEV CORP.  
 XX  
 PI Lindquist S, Queitsch C, Vierling E;  
 XX WPI; 2001-639123/73.  
 DR  
 PT Transgenic plants with improved heat stress tolerance, useful for  
 PT producing animal feed, oil and synthetic products.  
 XX  
 PS Claim 4; Page; 91pp; English.  
 XX



The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from *Streptococcus pneumoniae*. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the *S.pneumoniae* genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the *S. pneumoniae* genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the *S. pneumoniae* genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for *S. pneumoniae*

Seq	Sequence	11303 BP;	3733 A;	2266 C;	1558 G;	3746 T;	0 U;	0 Other;
	Query Match		10.4%;	Score 32.4;	DB 2;	Length 11303;		
	Best Local Similarity	51.4%;	Pred. No. 27;					
	Matches	75;	Conservative	0;	Mismatches	71;	Indels	0; Gaps 0;
QY	62	AAAGAGATTTCATCATCCAGAAAAGGTGTGGAAAAGTTGTGGATTCGGTGGAAAAAGTTC	121					
Db	6373	AGACAGAGTAAATTTACTTTGAAAAGATGGTTGAGAGTTATCTCTTCGGTGAATCAAAATGA	6314					
QY	122	GATCTGACCACTCTTAGATCGTGGAAAAAGTTTCACGTTAGCGCTTACGTACATATGTGGA	181					
Db	6313	GTTTGAACAATCTTTACTTTATGGAATTAACACGCGGGAAGATTAAGTTCAGTTGTTTT	6254					
QY	182	TTGTGGAAAAAGACGGAGGCATC	207					
Db	6253	TTCTGAAAAAATTGCAGTTCGCGCATC	6228					

Search completed: July 3, 2005, 02:38:02  
Job time : 481.353 secs

**This Page Blank (uspl.c)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 03:19:18 ; Search time 640.235 Seconds  
(without alignments)  
3056.566 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 gtcgaccatcatttgaagg.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues 12626748  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCRS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	37.4	12.0	2515	13	US-10-027-632-59048 Sequence 59048, A
C 2	37.4	12.0	2515	17	US-10-027-632-59048 Sequence 59048, A
C 3	36.8	11.8	350764	13	US-10-087-192-1864 Sequence 1864, Ap
C 4	33.8	10.8	1491	19	US-10-437-963-52281 Sequence 52281, A
C 5	33	10.6	4074	17	US-10-320-797-210 Sequence 210, App
C 6	32.8	10.5	657	13	US-10-027-632-259460 Sequence 259460,
C 7	32.8	10.5	657	17	US-10-027-632-259460 Sequence 259460,

8	32.8	10.5	145806	20	US-10-719-993-6943 Sequence 6943, Ap
9	32.6	10.4	468	9	US-09-974-300-3550 Sequence 3550, Ap
10	32.6	10.4	585	19	US-10-767-701-29171 Sequence 29171, A
11	32.6	10.4	1163020	17	US-10-398-221-10 Sequence 10, Appl
12	32.6	10.4	3011208	17	US-10-398-221-2058 Sequence 2058, Ap
13	32.4	10.4	495	21	US-10-472-928-445 Sequence 445, App
14	32.4	10.4	2992	9	US-09-812-350-43 Sequence 43, Appl
C 15	32.4	10.4	11303	8	US-08-961-527-115 Sequence 115, App
C 16	32.4	10.4	11303	17	US-10-158-844-115 Sequence 11, Appl
C 17	32.4	10.4	65454	18	US-10-293-864-11 Sequence 50, Appl
C 18	32.4	10.4	83517	19	US-10-367-094-50 Sequence 17645, A
C 19	32.4	10.4	103860	21	US-10-741-600-17645 Sequence 4790, Ap
C 20	32.4	10.4	2162598	21	US-10-472-928-4979 Sequence 4979, Ap
C 21	32.2	10.3	370	19	US-10-767-701-4790 Sequence 236927,
C 22	32.2	10.3	625	13	US-10-027-632-236927 Sequence 236928,
C 23	32.2	10.3	625	17	US-10-027-632-236927 Sequence 236927,
C 24	32.2	10.3	625	17	US-10-027-632-236927 Sequence 236928,
C 25	32.2	10.3	625	17	US-10-027-632-236927 Sequence 236927,
C 26	32	10.3	765	13	US-10-027-632-12112 Sequence 12112, A
C 27	32	10.3	765	13	US-10-027-632-12112 Sequence 12112, A
C 28	32	10.3	765	17	US-10-027-632-12112 Sequence 12112, A
C 29	32	10.3	765	17	US-10-027-632-12112 Sequence 12112, A
C 30	31.8	10.2	722	17	US-10-398-221-1116 Sequence 1116, Ap
C 31	31.8	10.2	2232	17	US-10-398-221-3638 Sequence 3638, Ap
C 32	31.8	10.2	3759	17	US-10-282-122A-16091 Sequence 16091, A
C 33	31.6	10.1	592	9	US-09-770-152-45 Sequence 45, Appl
C 34	31.6	10.1	2526	10	US-09-934-455-119 Sequence 119, App
C 35	31.6	10.1	2526	15	US-10-278-536-199 Sequence 199, App
C 36	31.6	10.1	2526	17	US-10-225-086A-793 Sequence 793, App
C 37	31.6	10.1	2526	17	US-10-374-780A-145 Sequence 145, App
C 38	31.6	10.1	2526	18	US-10-412-699B-463 Sequence 463, App
C 39	31.6	10.1	2526	18	US-10-412-699B-1805 Sequence 1805, Ap
C 40	31.4	10.1	534	17	US-10-257-166-96 Sequence 96, Appl
C 41	31.4	10.1	443	13	US-10-027-632-177537 Sequence 177537,
C 42	31.4	10.1	443	17	US-10-027-632-177537 Sequence 177537,
C 43	31.2	10.1	3673778	16	US-10-312-841-1 Sequence 2105, Ap
C 44	31.2	10.0	9859	15	US-10-311-455-2105 Sequence 1, Appl
C 45	31	9.9	783	18	US-10-424-599-97972 Sequence 97972, A

ALIGNMENTS

RESULT 1  
US-10-027-632-59048/c  
; Sequence 59048, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 10827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59048  
; LENGTH: 2515  
; TYPE: DNA

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2515)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59048

Query Match      12.0%; Score 37.4; DB 13; Length 2515;
Best Local Similarity 53.4%; Pred. No. 0.79;
Matches 77; Conservative 1; Mismatches 67; Indels 0; Gaps 0;

Qy 13 TTCAAGGGCCTCGGTAATACCACTTGTGGAAAAGTTGGTAATACGGAAGAAAGATTC 72
Db 1051 TTAAGGGCCTAGATTATGATATAGGAAGGGCTAAGGCAATGTTGATAAAGCATGTTA 992

Qy 73 ATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGGCTTGGAAAGTTGCATCTGACCAT 132
Db 991 ATGTGCTGGAAGCGTGTAAGAAAAGTAGTGTAGTTGATCATGAAGTGAGGCCCGGCAC 932

Qy 133 CTCTAGATCTGGAAAAGTTTCACG 157
Db 931 CCCAGGACAAAGGATGGAGTCCACG 907

RESULT 2
US-10-027-632-59048/c
; Sequence 59048, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59048
; LENGTH: 2515
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2515)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-59048

Query Match      12.0%; Score 37.4; DB 17; Length 2515;
Best Local Similarity 53.4%; Pred. No. 0.79;
Matches 77; Conservative 1; Mismatches 67; Indels 0; Gaps 0;

Qy 13 TTCAAGGGCCTCGGTAATACCACTTGTGGAAAAGTTGGTAATACGGAAGAAAGATTC 72
Db 1051 TTAAGGGCCTAGATTATGATATAGGAAGGGCTAAGGCAATGTTGATAAAGCATGTTA 992

Qy 73 ATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGGCTTGGAAAGTTGCATCTGACCAT 132
Db 991 ATGTGCTGGAAGCGTGTAAGAAAAGTAGTGTAGTTGATCATGAAGTGAGGCCCGGCAC 932
```

```
Qy 133 CTCTAGATCTGGAAAAGTTTCACG 157
Db 931 CCCAGGACAAAGGATGGAGTCCACG 907

RESULT 3
US-10-087-192-1864
; Sequence 1864, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1864
; LENGTH: 350764
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1864

Query Match      11.8%; Score 36.8; DB 13; Length 350764;
Best Local Similarity 56.7%; Pred. No. 11;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 70 TTCATCATCCAGAAAAGTTGTGGAAGTTGTGGATTGGTGGAAAAAGTTTCGATCTGAC 129
Db 176489 TTCTTTTAAAAAAATGATGCAAAAGCTGTTGCTTGGCGGGGTATTTTACATCAGAA 176548

Qy 130 CATCTCTAGATCTGGAAAAGTTACGTTAGCGCTTACATATGTGGATTGTGGAA 189
Db 176549 ATGCTTAAGAGATTGTAATAGGTTGGTGA AAAA AACTGAGGTACATTTAATGAATATGGAA 176608

RESULT 4
US-10-437-963-52281
; Sequence 52281, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 52281
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54592C.1
US-10-437-963-52281

Query Match      10.8%; Score 33.8; DB 19; Length 1491;
Best Local Similarity 53.4%; Pred. No. 8.3;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

QY 41 GAAAAAGTTGGTATACGGAAGAAAGAGATTTCATCCAGAAAGGTGTGAAAAAGTTG 100  
|||||  
Db 1126 GAAATTTGTGAAATTTGACAAAAGAGTTTCTTATTGAAAGATTCTTTGGAAAGTTT 1185  
|||||  
QY 101 TGGATTGCGTGGGAAAGTTGATCTGACCATCTCTAGATCGTGGAAAGTTTCAAGTTA 160  
|||||  
Db 1186 TCGATGGTCAAGAACCTCAACATGATCTTCTGATCAATCTAGCAAAATCAGCTTTT 1245  
|||||  
QY 161 GCGCTTACGTACA 173  
|||||  
Db 1246 GGTTCGAGGAACA 1258  
|||||

RESULT 5  
US-10-320-797-210/c  
; Sequence 210, Application US/10320797  
; Publication No. US20040014955A1  
; GENERAL INFORMATION:  
; APPLICANT: Eroskin, Alexey M.  
; APPLICANT: Zamudio, Carlos  
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 10182-021-999  
; CURRENT APPLICATION NUMBER: US/10/320,797  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: 60/341,261  
; PRIOR FILING DATE: 2001-12-17  
; NUMBER OF SEQ ID NOS: 3361  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 210  
; LENGTH: 4074  
; TYPE: DNA  
; ORGANISM: Cryptococcus neoformans  
US-10-320-797-210

Query Match 10.6%; Score 33; DB 17; Length 4074;  
Best Local Similarity 52.6%; Pred. No. 23;  
Matches 72; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 16 AAAGGCGCTCGGTAATACCATCTGTGAAAAAGTTGGTAATACGGAAGAAAGAGATTTCATC 75  
|||||  
Db 3865 AATGTGCCCAATGATATATTTGGTCAAGACACATATTCATCAAAAGCATTCTTT 3806  
|||||  
QY 76 ATCCAGAAAGGTGTGGAAGTTGTGGAATGTGGAATGTGGAAGTTGTGATCTGACCATCTC 135  
|||||  
Db 3805 TGTAAATAAGGCATGAGATGTTGAGGCTTTTGTAAACAAAGATTGAATTGAATGCTG 3746  
|||||  
QY 136 TAGATCGTGAAAAAGT 152  
|||||  
Db 3745 GGGATGGTGATACAAAT 3729  
|||||

RESULT 6  
US-10-027-632-259460/c  
; Sequence 259460, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 259460  
; LENGTH: 657  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-259460

Query Match 10.5%; Score 32.8; DB 13; Length 657;  
Best Local Similarity 56.5%; Pred. No. 12;  
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
QY 49 TGGTAATACGGAAGAAAGAGATTTCATCCAGAAAGGTGTGAAAAAGTTGTGGATTGC 108  
|||||  
Db 560 TGTAAAAAGGAACAAGAGGAAGATATATACAGAAACTCTGAAAAAGAGAACAAATGA 501  
|||||  
QY 109 GTGGAAAAAGTTGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCAC 156  
|||||  
Db 500 GAAGGATAAATCTTTTCAGATATTATTTAAACTTAGTTAAATATATAC 453  
|||||

RESULT 7  
US-10-027-632-259460/c  
; Sequence 259460, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 259460  
; LENGTH: 657  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-259460

Query Match 10.5%; Score 32.8; DB 17; Length 657;  
Best Local Similarity 56.5%; Pred. No. 12;  
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
QY 49 TGGTAATACGGAAGAAAGAGATTTCATCCAGAAAGGTGTGAAAAAGTTGTGGATTGC 108  
|||||  
Db 560 TGTAAAAAGGAACAAGAGGAAGATATATACAGAAACTCTGAAAAAGAGAACAAATGA 501  
|||||  
QY 109 GTGGAAAAAGTTGATCTGACCATCTCTAGATCGTGGAAAAAGTTTCAC 156  
|||||  
Db 500 GAAGGATAAATCTTTTCAGATATTATTTAAACTTAGTTAAATATATAC 453  
|||||

RESULT 8  
US-10-719-993-6943

; Sequence 6943, Application US/10719993  
; Publication No. US20040285849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6943  
; LENGTH: 145806  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-719-993-6943

Query Match 10.5%; Score 32.8; DB 20; Length 145806;  
Best Local Similarity 53.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Matches 70; Conservative 0; Mismatches 62;  
QY 8 ATCATTTGAAGGCGCTCGGTAATACCATTTGTGGAAGAGTTGGTAATACGGAAAAAGAA 67  
Db 82601 ATTCATTGCAAGGTGTGTAAGGACTAGAAAGCAAGAGGGAATACAGGTTCAAAA 82660  
QY 68 GATTCATCATCCAGAAAAGGTGGAAGAGTTGTGGATTCGCGGAAAAAGTTTCGATCTG 127  
Db 82661 GATTAATAATTCAGAAAGGTGTGTTGCTCTGGGATGTAGAAGCAAGAAAAAATA 82720  
QY 128 ACCATCTCTAGA 139  
Db 82721 ATATTACCCAGA 82732

RESULT 9  
US-09-974-300-3550  
; Sequence 3550, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3550  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-3550

Query Match 10.4%; Score 32.6; DB 9; Length 468;  
Best Local Similarity 55.9%; Pred. No. 12; Indels 0; Gaps 0;  
Matches 62; Conservative 0; Mismatches 49;  
QY 21 GCTTCGGTAATACCATTTGGAAAAAGTTGGTAATACGGAAAAAGAAAGATTCAATCA 80  
Db 191 GTCTCAACAGTCCGGCTTTACACAGATTCAAAATCGGGAAGGCGATTTTTTATCC 250  
QY 81 GAAAGGTGGAAAAAGTTGTGATTCGCGTGGAAAAAGTTTCGATCGACCA 131  
Db 251 GGAAGGGCTTGAACCTTTAGACAGTCCATGGAAAAAATGGGGAGCTGACAA 301

RESULT 10  
US-10-767-701-29171

; Sequence 29171, Application US/10767701  
; Publication No. US20040172864A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 29171  
; LENGTH: 585  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 9296737  
US-10-767-701-29171

Query Match 10.4%; Score 32.6; DB 19; Length 585;  
Best Local Similarity 55.9%; Pred. No. 13; Indels 0; Gaps 0;  
Matches 62; Conservative 0; Mismatches 49;  
QY 88 TCTGGAAGTTGTGGATTGCGTGAAGAAAGTTGATCTGACCATCTCTAGATCGTGAA 147  
Db 102 TGAATCAAGAGTATACGATGGTTGAAACGGCGCAACACAGGCTCGACTTCGGGAA 161  
QY 148 AAGATTTCACGTTAGCGCTTACGTACATATGTGGATTGTGGAAGAAAGAAC 198  
Db 162 AAGCTCGAGCTGAGCGCATGCGTAAAGCTATGCGGAGGTCAAGGAAGCC 212

RESULT 11  
US-10-398-221-10  
; Sequence 10, Application US/10398221  
; Publication No. US20040018514A1  
; GENERAL INFORMATION:  
; APPLICANT: KUNST, Frederik  
; APPLICANT: GLASER, Philippe  
; TITLE OF INVENTION: Listeria innocua, genome and applications  
; FILE REFERENCE: 344 702 - US  
; CURRENT APPLICATION NUMBER: US/10/398,221  
; CURRENT FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: FR 00/12 697  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4025  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1163020  
; TYPE: DNA  
; ORGANISM: Listeria innocua  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(end)  
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u  
US-10-398-221-10

Query Match 10.4%; Score 32.6; DB 17; Length 1163020;  
Best Local Similarity 51.7%; Pred. No. 3.6e+02; Indels 0; Gaps 0;  
Matches 74; Conservative 0; Mismatches 69;  
QY 62 AAAGAAGATTTCATCATCCAGAAAGGTGTGGAAGAGTTGTGGATTCGCGTGGAAAAAGTTTC 121  
Db 61761 AAATCAGGTTTATAAATCAAAATCGGGTGAAGAGATGAATAAGAAAGATGAAAGCTT 61820  
QY 122 GATCTGACCATCTCTAGATCGGAAAAAGTTTCAGTTAGCGCTTACGTACATATGTGGA 181  
Db 61821 TTTAGTATTACTTCTGTTGTTGTTAGTATTATTCGTTAGTGGTTGTTATCAATTGTTGCA 61880  
QY 182 TTGTGGAAGAAAGAACGAGGC 204

[illegible]

```
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-115

Query Match      10.4%; Score 32.4; DB 8; Length 11303;
Best Local Similarity 51.4%; Pred. No. 55;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      62 AAAGAAGATTTCATCATCCAGAAAAGGTGTGGAAGTTGTGGATTGCGTGGAAGAAAGTTC 121
Db      6373 AGACAAGGTAAATTACTTGAAAAGATGGTTGAGAGTTTATCTGTTGGTGAATCAATGA 6314

QY      122 GATCTGACCATCTCTAGATCGTGGAAAAAGTTCCACGTTAGCGCTTACGTACATATGTGGA 181
Db      6313 GTTTGAACATCTTTACTTTATGGAATTAACACAGCGGAAGATTAAAGTTCAGTTGTTTT 6254

QY      182 TTGTGGAAGAAAGACGGAGGCATC 207
Db      6253 TTCTGAAAAAATTGCAGTTCGCCATC 6228

Search completed: July 3, 2005, 07:33:13
Job time : 652.235 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 02:01:47 ; Search time 3925.41 Seconds  
(without alignments)  
3025.428 Million cell updates/sec

Title: US-10-814-858A-1  
Perfect score: 312  
Sequence: 1 gtcgaccatcattgaaag.....tgacgcacttgacgtact 312

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_hic.\*
- 4: gb\_est3.\*
- 5: gb\_est4.\*
- 6: gb\_est5.\*
- 7: gb\_est6.\*
- 8: gb\_gss1.\*
- 9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	12.4	777	8 BZ096570	CH230-236
2	38	12.2	1007	9 CNS00JOV	AL076752 Drosophil
3	37.8	12.1	1101	9 CNS00HBP	AL073282 Drosophil
4	37.2	11.9	754	8 BZ063129	BZ063129 l1d86503.
5	36.6	11.7	925	9 CNS00IBN	AL074475 Drosophil
6	36.4	11.7	765	9 CC807465	CC807465 ZMMB8C051
7	36.4	11.7	2707	3 AK048004	AK048004 Mus muscu
8	36.2	11.6	633	7 CF979725	CF979725 tg80c09.Y
9	36	11.5	368	8 BZ840661	BZ840661 CH240.261
10	36	11.5	576	9 CE32559	CE32559 tigr-gss-
11	36	11.5	960	9 CNS0062M	AL066226 Drosophil
12	36	11.5	1070	9 AG120841	AG120841 Pan trogl
13	35.8	11.5	367	8 BZ852956	BZ852956 CH240.217
14	35.6	11.4	921	9 AG561995	AG561995 Mus muscu
15	35.6	11.4	1101	9 CNS0106X	AL098595 Drosophil
16	35.4	11.3	507	9 CR498396	CR498396 Medicago
17	35.4	11.3	508	8 AQ439862	AQ439862 HS 5059.A
18	35.4	11.3	512	9 BX185081	BX185081 Danio rer
19	35.4	11.3	513	8 AZ570160	AZ570160 271PvB08
20	35.4	11.3	585	9 BX198554	BX198554 Danio rer
21	35.4	11.3	1449	9 AG391030	AG391030 Mus muscu
22	35.2	11.3	375	9 CE224908	CE224908 tigr-gss-
23	35.2	11.3	392	9 CE224907	CE224907 tigr-gss-
24	35.2	11.3	567	4 BI667582	BI667582 603292418

c	25	35.2	11.3	751	8	BZ074256	1kh06g12.
	26	35.2	11.3	797	9	CNS003P8	AL064634 Drosophil
	27	35.2	11.3	924	9	CNS0056T	AL057103 Drosophil
	28	35	11.2	1002	9	CL948189	OL1F8B003
	29	34.8	11.2	621	6	CB435223	CB435223 615287.MA
	30	34.8	11.2	642	4	BG538051	BG538051 602563558
	31	34.8	11.2	945	7	CF584573	CF584573 AGENCOURT
	32	34.6	11.1	1100	9	CNS00G3S	AL071398 Drosophil
	33	34.4	11.1	492	1	AV675381	AV675381 AV675381
	34	34.4	11.0	611	5	BW045972	BW045972 BW045972
	35	34.4	11.0	630	5	BW264718	BW264718 BW264718
	36	34.4	11.0	641	5	BW337262	BW337262 BW337262
	37	34.4	11.0	647	5	BW052958	BW052958 BW052958
	38	34.4	11.0	651	5	BW030523	BW030523 BW030523
	39	34.4	11.0	681	5	BW301819	BW301819 BW301819
	40	34.4	11.0	700	5	BW486169	BW486169 BW486169
	41	34.4	11.0	714	7	CF368490	CF368490 853144.MA
	42	34.4	11.0	755	5	BW299977	BW299977 BW299977
c	43	34.4	11.0	759	5	BW424685	BW424685 BW424685
	44	34.4	11.0	769	5	BW043106	BW043106 BW043106
	45	34.4	11.0	789	1	AL666171	AL666171 AL666171

ALIGNMENTS

RESULT 1  
BZ096570  
LOCUS BZ096570 777 bp DNA linear GSS 10-OCT-2002  
DEFINITION CH230-236C5-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-236C5, genomic survey sequence.  
ACCESSION BZ096570  
VERSION BZ096570.1 GI:23737454  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 777)  
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K.,  
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
Riggs,F., de Jong,P. and Fraser,C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)  
JOURNAL Other GSSs: CH230-236C5.TJ  
COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or ering information.htm). BAC end  
page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 236 row: C column: 5  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .777  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SsNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-236C5"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

```
CHORI-230 Rat (BN/SenHed/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match          12.4%; Score 38.8; DB 8; Length 777;
Best Local Similarity 51.1%; Pred. No. 1.8;
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 43 AAAAGTTGGTAATACGGAAGAAAGATTTCATCATCCAGAAAAGGTGTGGAAAAGTTGTG 102
    |||||
Db 193 AGAAGAAGCAAAATTTTATAATAAGAAATATTAAACAAAACAAATACAGAAATATAGAT 252
    |||||

QY 103 GATTGCGTGAAAAGTTTCGATCTGACCATCTCTAGATCTGTGGAAAAGTTTCAGTTAGC 162
    |||||
Db 253 AAATCAAACTCAAAATTCGATCTTAACAATAACACAGCAAGGAAAGAAAATCAATTAAGC 312
    |||||

QY 163 GCTTACGTCATATGTTGGATTGTGGAAAAGAAAGCGGAGGCATCGGTGGAAAAGAA 220
    |||||
Db 313 ACTTATGGGTAAAGTAGAATAAGAAATAGAAATGTAATGAAGGATGGAAAAGGA 370
    |||||

RESULT 2
CNS00JOV/c
LOCUS
DEFINITION
Drosophila melanogaster 1007 bp DNA linear GSS 03-JUN-1999
BACR38B02 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL076752.1 GI:4956128
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1007)
AUTHORS
TITLE
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1007
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR38B02"
/clone_lib="RPCI-98"
/notes="end : TET3"

ORIGIN
Query Match          12.2%; Score 38; DB 9; Length 1007;
Best Local Similarity 39.7%; Pred. No. 3.4;
Matches 50; Conservative 29; Mismatches 47; Indels 0; Gaps 0;

QY 35 ATTGTGGAAAAGTTGGTAATACGGAAGAAAGATTTCATCCAGAAAAGGTGTGGAA 94
    |||||
Db 654 AWDAAADAADAATRAATARAADATKAAAKAAAAAATAATKTGGTARAATAWAKTGRKKDA 595
    |||||
```

```
QY 95 AAGTTTGGATTGCGTGGAAAAGTTTCGATCTGACCATCTCTAGATCTGTGGAAAAGTTTC 154
    |||||
Db 594 AAKKTTTGTGTRKGGTGAATAWATTKGTGAGDARTNKGAGAAAKAARAAGGK 535
    |||||

QY 155 ACGTTA 160
    |||
Db 534 AXGGDA 529
    |||

RESULT 3
CNS00HBP
LOCUS
DEFINITION
Drosophila melanogaster 1101 bp DNA linear GSS 03-JUN-1999
BACR34A09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL073282.1 GI:4953062
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
TITLE
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR34A09"
/clone_lib="RPCI-98"
/notes="end : TET3"

ORIGIN
Query Match          12.1%; Score 37.8; DB 9; Length 1101;
Best Local Similarity 34.2%; Pred. No. 4;
Matches 96; Conservative 46; Mismatches 139; Indels 0; Gaps 0;

QY 15 GAAAGGCGCTCGGTAATACCATTTGTGGAAAAGTTGGTAATACGGAAGAAAGATTTCAT 74
    |||||
Db 741 GAAAGGTTTTGKMWTTATCAATGYASAAGAGGTGTTAKDAYCAAGAGGGWTTTHTMAR 800
    |||||

QY 75 CATCCAGAAAAGGTGTGGAAAAGTTGTGGATTGCGTGGAAAAGTTTCGATCTGACCATCT 134
    |||||
Db 801 TTWTHAHMGATGTGCAWMAWDDTTCTMWCARTCTGGGAAWAWATTTKGARTMTYCAAT 860
    |||||

QY 135 CTAGATCGTGGAAAAGTTTCACGTAGCGCTTACGTACATATGTGCGATTGTGGAAAAGA 194
    |||||
Db 861 TWTCCGKSGGGGMMWTGGRAGCCCCCTTTTAWAAAAAARAAGMTATAWAAMWA 920
    |||||

QY 195 AGACGGAGGCATCGGTGGAAAAGAAAGCTTGTACGCTGACGTACGTAGATGATAC 254
    |||||
```



Db 921 AAAAAAGWAGTTGATGGGAWAAWNTWTGTGTTKTKWTATGTGTCGCCCGGCTCTTGWTAT 980

Qy 255 ACGTGCACGGCTCCACTTGACGCAATTTGACGCAATGA 295

Db 981 GGGCTKAHAGAGTATCGGNAATAWAAGGGGKYMTSWG 1021

RESULT 4

BZ063129

LOCUS

DEFINITION 754 bp DNA linear GSS 10-OCT-2002

ACCESSION 11d86g03.b1 B.oleracea002 Brassica oleracea genomic, genomic survey

VERSION BZ063129

KEYWORDS BZ063129.1 GI:23673631

SOURCE GSS.

ORGANISM Brassica oleracea

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

AUTHORS 1 (bases 1 to 754)

TITLE Delehaunty, K., Fellw, G., Fulton, L., McCombie, W.R., Miner, T., Nash, W., Rabinowicz, P.D. and Wilson, R.K.

JOURNAL Whole genome shotgun reads from Brassica oleracea

COMMENT Unpublished (2002)

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Plate: 11d86 row: g column: 03

Seq primer: -21UPpof forward

Class: Shotgun

High quality sequence start: 16

High quality sequence stop: 551.

FEATURES

source

1. .754

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/db\_xref="taxon:3712"

/clone\_lib="B.oleracea002"

/notes="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T0100DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 11.9%; Score 37.2; DB 8; Length 754;

Best Local Similarity 53.4%; Pred. No. 5.5;

Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 59 GAAAAAGAGATTCATCATCCAGAAAGGTGCGAAAGTTGGATTGGTGGAAAAAG 118

Db 462 GAAAGGCGAGCGAAGTAGAAAAACATTCAGTCTACAAATTTGATTTCTTATATAAT 521

Qy 119 TTCATCTGACCATCTCTACATCGTGAAAAAGTTTCACGTTAGCGTTAGTACATATGT 178

Db 522 TTGTTGATATTTTTCATGTAAAGTGTGTTAAAGTATGCTTTAAACGGAACCAATATTT 581

Qy 179 GGATTTGGAAAAAGACGCGAGGC 204

Db 582 GGATTCACAAAAAAGGGATTATGC 607

RESULT 5

CNS001BN

LOCUS

DEFINITION 925 bp DNA linear GSS 03-JUN-1999

ACCESSION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR36N21 of RCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL074475

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .925

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone\_lib="BACR36N21"

/clone\_lib="RCI-98"

/note="end : T7"

ORIGIN

Query Match 11.7%; Score 36.6; DB 9; Length 925;

Best Local Similarity 15.3%; Pred. No. 8.6;

Matches 33; Conservative 97; Mismatches 86; Indels 0; Gaps 0;

Qy 19 GGGCTTCGTAATACCATTTGTGGAAAAAGTTGGTATATACGAAAAAGAGATTTCATC 78

Db 668 GCGMGGGGGVMKKKKKXMDAMKMKMKMKMKMKMKKKKKKKKKKKKKKKKKKKKK 727

Qy 79 CAGAAAAGTGTGAAAAAGTTGTGGATTTCGGTGGAAAAAGTTGACATCTCTCTAG 138

Db 728 RAAAKADKKKKKKAADAAAADDKAKKKKOHAMMMMMMMMTKTKTKKMMMTTMMMTK 787

Qy 139 ATCGTGGAAAAAGTTTCACGTTAGCGTTACGTACATATGTTGATTTCGAAAAAGAGAC 198

Db 788 KGAAGAARAKKKAANAKAAKAKKKAKKKGAKKKKKKKKKKKKKKKKKKKKKKA 847

Qy 199 GGAGGCATCGGTGGAAAAAGAGCTTTGTACGCTGA 234

Db 848 AKKKKKKADKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 883

RESULT 6

CC807465

LOCUS

DEFINITION 765 bp DNA linear GSS 14-JUL-2003

ACCESSION ZMMBBc0514M03f ZMMBBc Zea mays genomic clone ZMMBBc0514M03 5',

VERSION genomic survey sequence.

KEYWORDS CC807465

SOURCE CC807465.1 GI:32638048

ORGANISM GSS.

Ze mays

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 765)

AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,  
Rouzaud,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.  
TITLE Sequencing of the maize genome at PGIR (2003b)  
JOURNAL Unpublished (2003)  
COMMENT Contact: Bharti,A.K.  
Dr Joachim Messing's lab  
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers  
University  
190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
Tel: 732 445 3801  
Fax: 732 445 5735  
Email: bharti@waksman.rutgers.edu  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 401.  
Location/Qualifiers  
source  
1. .765  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ZMBBc0514M03"  
/lab\_host="E. coli DH10B"  
/clone\_lib="ZMBBc"  
/note="Vector: pTARBAC1.3; Site\_1: BamHI; Site\_2: BamHI"

ORIGIN  
Query Match 11.7%; Score 36.4; DB 9; Length 765;  
Best Local Similarity 50.6%; Pred. No. 9.5;  
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 81 GAAAGGTGTGGAAGTTGTGATTCGTCGCGGAAAAAGTTGATCTGACATCTCTAGAT 140  
|||||  
491 GAAGAGAGGCGGAGAAATTTGGGATCGAGAGGCTAAACCTAACTAGACTAGAACTACTC 350  
|||||

QY 141 CGTGGAAAAGTTCACGTAGCGCTTACGTACATATGTGGATGTGGAAAAAGAACGG 200  
|||||

Db 551 CTAAGATAAAATCGGAAGTAGAAGTTGATGTGATTCGATTTGATTACAAATCGGCCGT 610  
|||||

QY 201 AGGCATCGGTGGAAGAAAGCTTTGACGCTGACGCTGACGATAGATAGATAC 254  
|||||

Db 611 AGACCTCTTTATATAGAGAGGGGGCTGGACCTTTACAGCTGGATTC 664  
|||||

RESULT 7  
AK048004  
LOCUS  
DEFINITION 2707 bp mRNA linear HTC 03-APR-2004  
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched  
library, clone:Cl30029F03 product:unclassified, full insert  
sequence.  
AK048004  
VERSION AK048004.1 GI:26092584  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
1  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

REFERENCE  
3  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4  
THE RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE  
5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE  
6  
AUTHORS Adachi,T., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saichou,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,E., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
TITLE Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.jp,  
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.jp/  
URL:http://fantom.gsc.riken.jp/  
Location/Qualifiers  
source  
1. .2707  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strains="CS7BL/6J"  
/db\_xref="FANTOM,DB:Cl30029F03"  
/db\_xref="taxon:10090"  
/clone="Cl30029F03"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
misc\_feature  
1. .2707  
/note="unclassified"

ORIGIN  
Query Match 11.7%; Score 36.4; DB 3; Length 2707;  
Best Local Similarity 52.7%; Pred. No. 13;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 149 AAGTTTCACGTTAGCGCTTACGTACATATGTGATTTGTGAAAAAGACGAGGCGATCG 208  
|||||

Db 2145 AAATTCCTTTTCTTCTTACACATATGTGATGTGTGCACACACACGACACAGG 2204  
|||||

Qy 209 GTGAAAAGAAGCTTGTACGCTGACGATAGATAGATACACGTCACGCGTCC 268  
 |||||  
 Db 2205 GTGCATACATATGTGTATGTATACACACACACACACACACACACACAC 2264  
 |||||  
 Qy 269 ACTTGACGCACAATGACGACAAATGACGC 298  
 |||||  
 Db 2265 ACACACACACACTCACACACACACACAC 2294  
 |||||

RESULT 8  
 LOCUS CF979725 633 bp mRNA linear EST 25-NOV-2003  
 DEFINITION r80609.y1 Meloidogyne hapla female SL1 pGEM Meloidogyne hapla cDNA  
 5' similar to TR:Q5XW8 Q9XW8 Y37D8A.14 PROTEIN. [1] , mRNA  
 sequence.  
 ACCESSION CF979725  
 VERSION CF979725.1 GI:38512774  
 SOURCE EST.  
 ORGANISM Meloidogyne hapla  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
 1 (bases 1 to 633)  
 McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,  
 Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,  
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,  
 Taagareishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,  
 Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,  
 Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
 McCann, R., Waterston, R. and Wilson, R.  
 The Washington Univ. Nematode EST Project, 1999  
 Unpublished (1999)  
 CONTACT: McCarter JP

TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Washington University School of Medicine  
 COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 Cloned unidirectionally. Poly (A)+ RNA was concentrated and  
 purified using Dynabeads (Dyna) and mRNA eluted for 1st strand  
 synthesis. 1st strand cDNA created using MMLV RT(PowerScript,  
 Clontech) and primed with oligo(dT) with XhoI site (primer  
 CDSIII/3-XhoI) and 5' SMART anchor added using chimeric DNA-RNA  
 oligo (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a PCR  
 using the SMART template was also desired. 15 PCR cycles were done  
 using 1st strand and primers specific to SL1 leader sequence  
 (SL1-NotI) and 3' end (XhoI-No-dT). dscDNA was digested using  
 XhoI/NotI, fractionated on Chroma-spin 400 columns (Clontech) and  
 ligated to digested pGEM-11zf(+) plasmid. Chemically competent  
 cells were used as host. This library was provided by David Bird's  
 lab (david.bird@ncsu.edu), of North Carolina State University.  
 Seq primer: Sp6.

FEATURES  
 source Location/Qualifiers  
 1. .633  
 /organism="Meloidogyne hapla"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6305"  
 /sex="female"  
 /tissue\_type="whole organism"  
 /dev\_stage="adult"  
 /lab\_host="DH108"  
 /clone\_lib="Meloidogyne hapla female SL1 pGEM"  
 /note="Vector: pGEM-11zf(+). Promega; Site 1: XhoI;  
 Site 2: NotI; Cloned unidirectionally. Poly (A)+ RNA was  
 concentrated and purified using Dynabeads (Dyna) and mRNA  
 eluted for 1st strand synthesis. 1st strand cDNA created  
 using MMLV RT(PowerScript, Clontech) and primed with  
 oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and 5'  
 SMART anchor added using chimeric DNA-RNA oligo  
 (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a  
 PCR using the SMART template was also desired. 15 PCR

cycles were done using 1st strand and primers specific to  
 SL1 leader sequence (SL1-NotI) and 3' end (XhoI-No-dT).  
 dscDNA was digested using XhoI/NotI, fractionated on  
 Chroma-spin 400 columns (Clontech) and ligated to digested  
 pGEM-11zf(+) plasmid. Chemically competent cells were  
 used as host. This library was provided by David Bird's  
 lab (david.bird@ncsu.edu), of North Carolina State  
 University."

ORIGIN  
 Query Match 11.8%; Score 36.2; DB 7; Length 633;  
 Best Local Similarity 49.7%; Pred. No. 10;  
 Matches 92; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
 Qy 11 ATTGGAAGGGCTCCGTAATACCATTTGCGAAAAAGTTGGTAATACGGAAAAAGAGAT 70  
 |||||  
 Db 449 ATGACAAATGCGCTGGATATGAACATTTTGCATAGAGTTGTAGAGAAATAAATAGAA 508  
 |||||  
 Qy 71 TCATCATCCAGAAAAGGTGTGGAATAATTTGTGATTCGCTGGAAAAAGTTCTGATCTGACC 130  
 |||||  
 Db 509 AGAAGTGGTAATATTTGTAGGGATATTTCTAAATGATTTAAATAAATTAATTTTCC 568  
 |||||  
 Qy 131 ATCTTAGATCGTGGAAAAAGTTTCACGTTAGCCCTACGTACATATGTTGGATTTGTGAAA 190  
 |||||  
 Db 569 CTCTTTAAAGTAAAAAATAATTTTGAAGAATTTCTGATTAATAATTTATTTTTTTTAAAA 628  
 |||||  
 Qy 191 AAGAA 195  
 |||||  
 Db 629 AAAAA 633

RESULT 9  
 LOCUS BZ840661 368 bp DNA linear GSS 18-MAR-2003  
 DEFINITION CH240\_261B20.TJ CHORI-240 Bos taurus genomic clone CH240\_261B20,  
 genomic survey sequence.  
 ACCESSION BZ840661  
 VERSION BZ840661.1 GI:29068020  
 KEYWORDS GSS.  
 SOURCE Bos taurus (cow)  
 ORGANISM

REFERENCE 1 (bases 1 to 368)  
 AUTHORS Zhao, S., Shetty, J., Shateman, S., Tsengaye, G., Geer, K.,  
 Shvartsbeyn, A., Gebregorgis, E., Chen, D., Riggs, F., de Jong, P.,  
 Crawford, A. M. and McEwan, J. C.  
 Bovine BAC End Sequences from Library CHORI-240  
 Unpublished (2003)

TITLE Contact: Shaying Zhao  
 JOURNAL Department of Eukaryotic Genomics  
 COMMENT The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@igr.org  
 Clones are derived from the bovine BAC library CHORI-240  
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/ordering/information.htm). This work  
 was undertaken as part of the International Bovine BAC Mapping  
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and The  
 Institute of Genomic Research (TIGR), USA.  
 Seq primer: Sp6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1. .368  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="breed: Hereford"



```

Db          793 GGNMMGGMVGNMG 779

RESULT 12
AG120841/c
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-129L09.F, genomic survey sequence.
ACCESSION AG120841
VERSION AG120841.1 GI:16650006
KEYWORDS GSS.
SOURCE
ORGANISM Pan troglodytes (chimpanzee)

REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111 Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
1. .1070
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-129L09.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match 11.5%; Score 36; DB 9; Length 1070;
Best Local Similarity 48.5%; Pred. No. 13;
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 44 AAAGTTGGTAATACGGAAGAAAGAGATTCATCCGAAAGAGGTGTGGAAGTTGTGG 103
Db 1070 AAATAAGGGAAGGGGAAAGCGGAGGAATAGGAGAAAGAGGAGGAGATAGATGGGA 1011

QY 104 ATTGCGTGGAAAAGTTTCGATCTGACCATCTCTAGATCGGGAAGAGTTTCAGTTAGCG 163
Db 1010 ATAGGAGAGAAAGAGAGATAGAGATATAATAAAGAGGAGGAGGAGGGAGGT 951

QY 164 CTTACGTACATATGTGGATTGTGGAAAAGAGACGAGGAGCATCGGTGGAAAAGAGAGCT 223
Db 950 CGAAAGAGAGAGGTGTGAGAGGAGAGAAAGAAAGAGAGAGTGAAGGAGGAGGAGGA 891

QY 224 TGTACGCTGTACGCTGACGATAGA 247
Db 890 GGGAGGGGACAGAAAGAGAGAGA 867

RESULT 13
BZ852956
LOCUS
DEFINITION BZ852956
VERSION BZ852956.1 GI:29080361
KEYWORDS GSS.
SOURCE
ORGANISM Bos taurus (cow)

REFERENCE
AUTHORS Zhao,S., Shetty,J., Shateman,S., Teegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
Crawford,A.M. and McSwan,J.C.
TITLE Bovine BAC End Sequences from Library CHORI-240
JOURNAL Unpublished (2003)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orderinginformation.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
Institute of Genomic Research (TIGR), USA.
Plate: 217 row: K column: 24
Seq primer: T7
Class: BAC ends.
FEATURES
source
1. .367
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_217K24"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="Vector: PTARBAC1.3; Site 1: Mbol; Site 2: Mbol;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 11.5%; Score 35.8; DB 8; Length 367;
Best Local Similarity 50.9%; Pred. No. 12;
Matches 85; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 5 ACATCATTTGAAGGGCCCTCGGTAAATACCATTTGTGGAAGAAAGTTGGTAAATACGGA 64
Db 181 ACAAAAAATTTAAAACTGCAGCTTCATATCATTTGAAAAACAAGAAACTGAACATAAAA 240

QY 65 GAAGATTTCATCCAGAAAAGGTGTGGAAGAGTTTGGATTTCGGTGGAAAAGTTTCGAT 124
Db 241 GATGAAAAAATATCCAAAAGGCAAAATATGAAATATATGCTATTAAATGAAGACATTATTA 300

QY 125 CTGACCATCTCTAGATCGTGGAAAAGTTTCAGCTTAGCGCTTACGTA 171
Db 301 CAGGACATGTTTACATTATAGAAAACCTTGAACCTCTATTGTAATTA 347

RESULT 14
AG561995
LOCUS
DEFINITION AG561995
ACCESSION AG561995
VERSION AG561995.1 GI:48322693
KEYWORDS GSS.
Mus musculus molossinus DNA, clone:MSMg01-481K04.T7, genomic survey
sequence.

```

```

SOURCE      Mus musculus molossinus
ORGANISM    Mus musculus molossinus
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL    1
AUTHORS     Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
COMMENT     BAC end Sequences of Library MSMg01
JOURNAL    2 (bases 1 to 921)
AUTHORS     Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL    Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel:81-45-503-9111, Fax:81-45-503-9170
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
Library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Taukuba Institute Bio Resource Center, Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY     Vector      : pBACe3.6
R.Site 1   : EcoRI
R.Site 2   : EcoRI.
FEATURES    Location/Qualifiers
             1..921
             /organism="Mus musculus molossinus"
             /mol_type="genomic DNA"
             /sub_species="molossinus"
             /db_xref="taxon:57486"
             /clone="MSMg01-481K04.T7"
             /sex="male"
             /tissue_type="mixture of kidney and spleen"
             /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
Query Match      11.4%; Score 35.6; DB 9; Length 921;
Best Local Similarity 49.4%; Pred. No. 17;
Matches 83; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 53 AATACGGAAGAAAGATTTCATCATCCAGAAAGGTGTGGAAAAGTTGTGGATTGCGTGG 112
Db 472 AAAAAAGAAAAGGGGNAAAAAAAAAAAAAAAAAAGGGGAGAGAACCGGGGGCGGGAG 531

QY 113 AAAAAGTTCGATCGACCATCTAGATCGTGGAAAAGTTTCAGTTAGCGCTTAGCTAC 172
Db 532 AAAAAAAAAAAAAAAAAAAANCNNGAAAAAANAAAAAAAAAAAAAAAAAGCGCCGCGAG 591

QY 173 ATATGTGGATTGTGGAAGAAAGAGACGGAGGCATCGGTGGAAGAAAGAA 220
Db 592 AGCGGGGGGGGGGGAATAATATAGGGGGGGGGGAGAAAAA 639

RESULT 15
CNS0106X      1101 bp DNA linear GSS 26-JUL-1999
LOCUS        Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION   BHCN03K20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1 GI:5610206
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)

```

```

Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOAC11.
FEATURES    Location/Qualifiers
             1..1101
             /organism="Drosophila melanogaster"
             /mol_type="genomic DNA"
             /db_xref="taxon:7227"
             /clone="BACN03K20"
             /clone_lib="DrosBAC"
             /plasmid="pBelOAC11"
             /note="end : T7"
ORIGIN
Query Match      11.4%; Score 35.6; DB 9; Length 1101;
Best Local Similarity 15.7%; Pred. No. 18;
Matches 34; Conservative 106; Mismatches 76; Indels 0; Gaps 0;

QY 36 TTGTGGAAGAAAGTTGGTAATACGAAAAGAAAGATTTCATCCAGAAAAGGTGTGAAA 95
Db 818 KKKKDKAWGDKKAKKDKAAKAKADAAGDAKAKRRRAGDKDKAKADAKAKAKAKOD 877

QY 96 AGTTGTGGATTTCGTTGGAAAAGTTTCGATCTGACCATCTCTAGATCGTGGAAAAGTTCA 155
Db 878 DDAKAATKAKKATKAKDKAKAKKKKKDKDKAKAKAKADAKDKDDDKDKDKAKA 937

QY 156 CGTTAGCGCTTACGTACATATGTGGATTGTGGAAGAAAGACGGAGGCATCGGTGAAA 215
Db 938 DKKKWKORAKDKAKKDDKDKDKDKDKDKDKADAKAKAKADAKAKAKAKAKAAD 997

QY 216 AAGAAGCTTGTACGCTGTACGCTGACGATAGATAGATA 251
Db 998 DDAADAKAKADAKAKAKAKAKADDAKAKAKAKAKAKAKA 1033

Search completed: July 3, 2005, 04:51:59
Job time : 3933.41 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2005, 00:53:27 ; Search time 197.647 Seconds  
(without alignments)  
3893.639 Million cell updates/sec

Title: US-10-814-858A-2  
Perfect score: 130  
Sequence: 1 tctactatataggaagttc.....ttacatctagataaacaatg 130

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	100.0	452	3	AAA28449
2	89	68.5	89	3	AAA28445
3	63.6	48.9	3212	5	AAF30968
4	63.6	48.9	3288	5	AAF30970
5	63.6	48.9	3329	5	AAF30969
6	61	46.9	1006	2	AAx81436
7	60.4	46.5	1138	2	AAx81436
8	60.4	46.5	1138	2	AAx81436
9	60.4	46.5	1138	2	AAx81436
10	59.6	45.8	1138	2	AAx81436
11	58.2	44.8	2200	6	ABL58982
12	56.2	43.2	5033	2	AAQ74695
13	52	40.0	6395	2	AAQ95155
14	52	40.0	6395	2	AAQ95155
15	52	40.0	6395	2	AAQ95155
16	52	40.0	6395	2	AAQ95155
17	52	40.0	6425	4	AAx82330
18	52	40.0	6425	4	AAx82330
19	52	40.0	6439	4	AAx82330
20	52	40.0	6439	4	AAx82330

21	52	40.0	6446	2	AAZ20646	Aaz20646 TMV-based
22	52	40.0	6446	4	AAZ20646	Aaf82334 Tobacco m
23	52	40.0	6475	2	AAZ20645	Aaz20645 TMV-based
24	52	40.0	6475	4	AAZ20645	Aaf82333 Tobacco m
25	52	40.0	7926	3	AAZ89190	Aaz89190 Plasmid T
26	52	40.0	10600	11	ADM68432	Adm68432 Celery CE
27	52	40.0	10600	12	ADP26585	Adp26585 TMV CEL I
28	52	40.0	10600	13	ADQ88407	Adq88407 p117MP4-
29	52	40.0	10600	13	ADQ88407	Adq88407 p117MP4-
30	52	40.0	10624	11	ADM68433	Adm68433 Celery CE
31	52	40.0	10624	12	ADP26586	Adp26586 TMV CEL I
32	52	40.0	10624	13	ADQ88408	Adq88408 p117MP4-
33	52	40.0	10624	13	ADQ88408	Adq88408 p117MP4-
34	52	40.0	11222	12	ADN97500	Adn97500 Artificia
35	52	40.0	11641	6	AAx45229	AAx45229 TTODA (rg
36	52	40.0	11641	6	AAx45229	AAx45229 TTODA (rg
37	52	40.0	11641	10	ADB84772	ADB84772 Tobacco m
38	52	40.0	11641	12	ADU88298	Adu88298 TTODA chl
39	52	40.0	11641	12	ADM48702	Adm48702 TMV based
40	51.8	39.8	1825	2	AAQ67663	Aaq67663 TMV repli
41	51.8	39.8	1825	2	AAV54825	Aav54825 Replicon
42	51.8	39.8	7688	5	AAx20293	Aax20293 Tobacco m
43	51.8	39.8	7688	6	AAx24477	Aax24477 Tobacco m
44	51.8	39.8	7688	8	ADA15016	Ada15016 Tobacco m
45	51.8	39.8	7688	10	ADB83324	ADB83324 DNA trans

ALIGNMENTS

RESULT 1  
AAA28449  
ID AAA28449 standard; DNA; 452 BP.  
XX  
AC AAA28449;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Artificial synthetic promoter for high level plant gene expression.  
XX  
KW Artificial promoter; strong; weak; transgene expression; plant; ss.  
XX  
OS Synthetic.  
XX  
FH Key  
FT misc\_signal 1..120 Location/Qualifiers  
FT misc\_signal 121..142 /tag= a  
FT misc\_signal 143..220 /notes= "domain II"  
FT misc\_signal 221..266 /tag= b  
FT misc\_signal 267..307 /notes= "domain III"  
FT misc\_signal 308..324 /tag= c  
FT misc\_signal 325..351 /notes= "Minimal domain Ib"  
FT misc\_signal 352..452 /tag= d  
FT TATA\_signal 353..374 /notes= "Minimal domain Ia"  
FT TATA\_signal 375..452 /tag= e  
FT 5'UTR 453..452 /notes= "plant consensus TATA signal"  
XX  
EP1002869-A1.  
PD 24-MAY-2000.  
XX



```

PF 25-FEB-1999; 99EP-00301419.
XX
PR 09-NOV-1998; 98IN-DE003322.
XX
XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
PI Tuli R, Sawant SV, Singh PK, Gupta SK;
XX
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
XX expression of transgenes in different organisms.
XX
XX Claim 13; Page 11; 40pp; English.
XX
XX Chemically synthesized artificial promoters are new and comprise a DNA
XX sequence designed for a targeted level and pattern of gene expression by
XX strategically putting together several signature sequences identified by
XX sequence alignment and statistical analysis of a large database
XX constructed for this purpose. A method for chemically synthesizing an
XX artificial promoter for expressing genes at a desired level in different
XX organisms is also claimed. The high level expression in a plant using
XX such an artificial promoter (e.g. AAA28449) can be measured comprising
XX polyethylene glycol (PEG) mediated transformation of plant protoplasts as
XX well as biolistic mediated transformation of plant tissues including
XX root, stem, intact leaf tissue followed by transient GUS assay to compare
XX with a natural CaMV 35S promoter showing the desired level of activity.
XX The promoter is useful for high level expression of transgenes in
XX different organisms and for testing high level gene expression in plants
XX (claimed). The promoter is biologically active and is efficient and can
XX be synthesized to express in even the most complex organisms
XX
XX SQ Sequence 452 BP; 157 A; 88 C; 95 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 130; DB 3; Length 452;
XX Best Local Similarity 100.0%; Pred. No. 8.5e-22;
XX Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TCACATATATAGGAAGTTTCATTTCATTTGGAAATGACACGTTGTTCATTTCTCAACAA 60
Db 313 TCACATATATAGGAAGTTTCATTTCATTTGGAAATGACACGTTGTTCATTTCTCAACAA 372
XX
QY 61 TTACCAACACACACACACACACACACACATTTACATTTTACATTTACATCTAG 120
Db 373 TTACCAACACACACACACACACACACACATTTACATTTTACATTTACATCTAG 432
XX
QY 121 ATAAACAATG 130
Db 433 ATAAACAATG 442
XX
RESULT 2
AAA28445
ID AAA28445 standard; DNA; 89 BP.
XX
AC AAA28445;
XX
XX 29-AUG-2000 (first entry)
XX
XX Synthetic promoter 5' untranslated leader sequence.
XX
XX Artificial promoter; 5' untranslated leader sequence;
XX transgene expression; plant; ss.
XX
XX Synthetic.
XX
XX EP1002869-A1.
XX
XX 24-MAY-2000.
XX
XX 25-FEB-1999; 99EP-00301419.
XX
XX 09-NOV-1998; 98IN-DE003322.
XX

```

```

XX (COUL ) CSIR COUNCIL SCI IND RES.
XX
XX Tuli R, Sawant SV, Singh PK, Gupta SK;
XX
XX WPI; 2000-341712/30.
XX
XX New chemically synthesized artificial promoter, useful high level
XX expression of transgenes in different organisms.
XX
XX Claim 9; Page 23; 40pp; English.
XX
XX This comprises a 5' untranslated leader (UL) sequence (+1 to +89) which
XX contains a translational enhancer "CAA" type region. The 5' UL of highly
XX expressed genes varies from 75-90 nucleotides, while that of the lowly
XX expressed genes varies from 100-300 nucleotides. Chemically synthesized
XX artificial promoters are new and comprise a DNA sequence designed for a
XX targeted level and pattern of gene expression by strategically putting
XX together several signature sequences identified by sequence alignment and
XX statistical analysis of a large database constructed for this purpose. A
XX method for chemically synthesizing an artificial promoter for expressing
XX genes at a desired level in different organisms is also claimed. The high
XX level expression in a plant using such an artificial promoter (e.g.
XX AAA28449) can be measured comprising polyethylene glycol (PEG) mediated
XX transformation of plant protoplasts as well as biolistic mediated
XX transformation of plant tissues including root, stem, intact leaf tissue
XX followed by transient GUS assay to compare with a natural CaMV 35S
XX promoter showing the desired level of activity. The promoter is useful
XX for high level expression of transgenes in different organisms and for
XX testing high level gene expression in plants (claimed). The promoter is
XX biologically active and is efficient and can be synthesized to express in
XX even the most complex organisms
XX
XX SQ Sequence 89 BP; 41 A; 20 C; 4 G; 24 T; 0 U; 0 Other;
XX
XX Query Match 68.5%; Score 89; DB 3; Length 89;
XX Best Local Similarity 100.0%; Pred. No. 4.3e-12;
XX Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 39 ACGTGTGTGTCATTTCTCAACATTTACCAACACACACACACACATTTATCAAT 98
Db 1 ACGTGTGTGTCATTTCTCAACATTTACCAACACACACACACACATTTATCAAT 60
XX
QY 99 TACTATTACATTTACATCTAGATAACA 127
Db 61 TACTATTACATTTACATCTAGATAACA 89
XX
RESULT 3
AAF30968
ID AAF30968 standard; cDNA; 3212 BP.
XX
XX AAF30968;
XX
XX 23-JUL-2001 (first entry)
XX
XX Aspergillus niger beta-glucosidase BGL1 cDNA in plasmid pJDB1.
XX
XX Beta-glucosidase; BGL1; aroma; alcoholic beverage; ethanol;
XX transgenic plant; tobacco; pJDB1; ss.
XX
XX Aspergillus niger.
XX
XX Key Location/Qualifiers
XX mat_peptide 391..2916
XX /tag= a
XX
XX WO200136586-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-IL000758.
XX

```



```

FT sig_peptide 391..507
FT /tag= b
FT /note= "Cell signal peptide"
FT mat_peptide 508..3039
FT /tag= c
FT /note= "encodes BGL1-HDEL fusion"
XX
XX WO200136586-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-IL000758.
XX
XX 19-NOV-1999; 99US-00443338.
XX
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shu W, Marton I, Siegel DL, Ben-Ami B, Dekel M, Shoseyov O;
XX
XX WPI; 2001-355624/37.
XX
XX P-PSDB; AAB82328.
XX
XX New nucleic acid encoding beta-galactosidase from Aspergillus niger,
XX useful e.g. for increasing aroma compound formation during alcoholic
XX fermentation.
XX
XX Disclosure; Page 101-103; 106pp; English.
XX
XX The present sequence is that of a cDNA construct in plasmid pJDCB1T
XX encoding a polypeptide (see AAB82328) comprising the Cell signal peptide,
XX followed by the mature region of Aspergillus niger B1 (GMI CC 324626)
XX beta-glucosidase BGL1 and the endoplasmic reticulum (ER) retaining
XX peptide HDEL. The cDNA is placed under the control of the 35S promoter
XX and nos terminator. The gene cassette in pJDCB1T was inserted into binary
XX vector pBINplus.. Disarmed Agrobacterium tumefaciens LB4404 was
XX transformed with the resulting vector and used to transform tobacco cv.
XX Samson plants. The gene expression cassette provided for accumulation of
XX BGL1 in the ER. BGL1 polynucleotides, polypeptides, host cells and
XX transgenic plants are useful for fermentation, particularly to increase
XX the level of aroma compounds in alcoholic drinks, as well as other
XX fermentation products of plant material, hydrolysing cellobiose and thus
XX increasing the level of fermentable glucose, increasing the production of
XX alcohol, such as ethanol, from plant material, increasing the aroma
XX released from a plant or a plant product, and increasing the hydrolysis
XX or transglycosylation of glycosides. Unlike previously known beta-
XX glucosidases, BGL1 is very stable over a wide range of temperatures and
XX pH, e.g. 68% retention of activity after 1 hour at 60 degree C
XX
XX Sequence 3288 BP; 782 A; 873 C; 876 G; 757 T; 0 U; 0 Other;
XX
XX Query Match 48.9%; Score 63.6; DB 5; Length 3288;
XX Best Local Similarity 87.7%; Pred. No. 6.2e-06;
XX Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2
XX
XX QY 12 AGGAAGTTCATTTCATTGGAAATGGACA--CGTGTCTGTCATTTCTCAACAATTACCAACA 69
XX |||||
XX Db 274 AGGAAGTTCATTTCATTGGAGAGGACAGCTCTTGAGATCCTTCAACATTACCAACA 333
XX |||||
XX QY 70 ACAACAAACAAACAAACAACTATTACAAATTACTATTACAAATTACA 115
XX |||||
XX Db 334 ACAACAAACAAACAAACAACT--TACAAATTACTATTACAAATTACA 377
XX |||||
XX
XX RESULT 5
XX AAF30969.
XX ID AAF30969 standard; cDNA; 3329 BP.
XX
XX AC AAF30969;
XX
XX DT 23-JUL-2001 (first entry)
XX
XX Aspergillus niger beta-glucosidase BGL1 cDNA in plasmid pJDCB1.
XX

```

KW Beta-glucosidase; BGL1; aroma; alcoholic beverage; ethanol;  
 KW transgenic plant; tobacco; pUDCB1; ss.  
 XX  
 OS Aspergillus niger.  
 OS Unidentified.  
 OS Chimeric.

XX Key Location/Qualifiers  
 FT CDS 391..3033  
 FT /\*tag= a  
 FT sig\_peptide 391..507  
 FT /\*tag= b  
 FT /\*note= "Cell signal peptide"  
 FT mat\_peptide 508..3030  
 FT /\*tag= c  
 FT /\*note= "BGL1 mature polypeptide"

XX WO200136586-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-IL000758.

XX 19-NOV-1999; 99US-00443338.

XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

XX Shu W, Marton I, Siegel DL, Ben-Ami B, Dekel M, Shoseyov O;

XX WPI: 2001-355624/37.

XX P-PSDB; AAB82327.

XX New nucleic acid encoding beta-galactosidase from Aspergillus niger,  
 PT useful e.g. for increasing aroma compound formation during alcoholic  
 PT fermentation.

XX Disclosure; Page 97-99; 106pp; English.

XX The present sequence is that of a cDNA construct in plasmid pUDCB1  
 CC encoding a polypeptide (see AAB82327) comprising the Cell signal peptide  
 CC and the mature region of Aspergillus niger B1 (CWI CC 324626) beta-  
 CC glucosidase BGL1. The cDNA is placed under the control of the 35S  
 CC promoter and nos terminator. The gene cassette in pUDCB1 was inserted  
 CC into binary vector pBINplus. Disarmed Agrobacterium tumefaciens LB4404  
 CC was transformed with the resulting vector and used to transform tobacco  
 CC cv. Samson plants. The gene expression cassette provided for secretion of  
 CC BGL1 into the apoplast. BGL1 polynucleotides, polypeptides, host cells  
 CC and transgenic plants are useful for fermentation, particularly to  
 CC increase the level of aroma compounds in alcoholic drinks, as well as  
 CC other fermentation products of plant material, hydrolysing cellobiose and  
 CC thus increasing the level of fermentable glucose, increasing the  
 CC production of alcohol, such as ethanol, from plant material, increasing  
 CC the aroma released from a plant or a plant product, and increasing the  
 CC hydrolysis or transglycosylation of glycosides. Unlike previously known  
 CC beta-glucosidases, BGL1 is very stable over a wide range of temperatures  
 CC and pH, e.g. 68% retention of activity after 1 hour at 60 degree C

XX Sequence 3329 BP; 790 A; 882 C; 883 G; 774 T; 0 U; 0 Other;

Query Match 48.9%; Score 63.6; DB 5; Length 3329;  
 Best Local Similarity 87.7%; Pred. No. 6.2e-06;  
 Matches 93; Conservative 0; Mismatches 9; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAAATGGACA--CGTGTGTGTCATTTCTCAACAATTACCAACA 69  
 Db 274 AGGAAGTTCATTTCATTTGGAGAGGACAGCTTCTTGAGATCCTTCAACAATTACCAACA 333

QY 70 ACAACAACACAAACAACATTTACATTTACTATTTCATTTACATTTACA 115

Db 334 ACAACAACACAAACAACAT--TACAATTACTATTTCATTTACATTTACA 377

RESULT 6

AAx81436

ID AAX81436 standard; DNA; 1006 BP.

XX

AC AAX81436;

XX

DT 25-AUG-1999 (first entry)

XX

DE Arabidopsis thaliana promoter fragment produced by PCR.

XX

KW plant promoter; minimal promoter; transcription activating element;  
 KW gene expression; transgenic plant; ss.

XX

OS Synthetic.

OS Arabidopsis thaliana.

XX

PN WO9931258-A1.

XX

PD 24-JUN-1999.

XX

PF 10-DEC-1998; 98WO-EP008162.

XX

PR 12-DEC-1997; 97EP-00203912.

XX

PA (MOGE-) MOGEN INT NV.

XX

PI Stuiver MH, Sijbolts FH;

XX

DR WPI; 1999-395188/33.

XX

PT Novel constitutive plant promoters.

XX

PS Example 10; Page 36-37; 44pp; English.

XX

CC The specification describes plant promoters, where elements from one  
 CC promoter which are responsible for specific expression are combined with  
 CC elements from a second promoter which are responsible for a complementary  
 CC expression pattern. The novel promoter then shows expression in the  
 CC tissues and developmental stages which form the expression pattern of  
 CC both promoters. The plant promoter is characterized in that it comprises  
 CC a minimal promoter and transcription activating elements from a set of  
 CC promoters, the elements having a complementary pattern and level of  
 CC transcription in a plant. The promoters of the invention are used to  
 CC express genes in transgenic plants. The present sequence was used in the  
 CC course of the invention

XX Sequence 1006 BP; 320 A; 182 C; 183 G; 321 T; 0 U; 0 Other;

Query Match 46.9%; Score 61; DB 2; Length 1006;  
 Best Local Similarity 88.3%; Pred. No. 2.4e-05;  
 Matches 91; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAAATGGACACGTCGTGTGTCATTTCTCAACAATTACCAACAAC 71

Db 906 AGGAAGTTCATTTCATTTGGAGAGGACAGGTAT----TTTTCACAACAATTACCAACAAC 960

QY 72 AACAAACAACAACAACATTTACATTTACTATTTCATTTACATTTAC 114

Db 961 AACAAACAACAACAACAT--TACAATTACTATTTCATTTACATTTAC 1001

RESULT 7

AAT85665

ID AAT85665 standard; DNA; 1138 BP.

XX

AC AAT85665;

XX

DT 21-NOV-1997 (first entry)

XX

DE pmJB1 DNA sequence.

XX

KW expression cassette; inducible promoter; alcA; alcR; aldA; regulator;  
 KW alcohol dehydrogenase; herbicide resistance gene; glyphosate;  
 KW N-phosphonomethyl-glycine; glyphosate oxidase; GOX; EPSPS;

KW 5-enol-pyruvylshikimate 3-phosphate CP4; chloroplast transit sequence 1;  
 KW Arabidopsis; RUBISCO; CPT1; dicotyledon; monocotyledon; crop; ss.  
 XX Synthetic.  
 OS Synthetic.  
 XX WO9706269-A1.  
 PN 20-FEB-1997.  
 XX 02-AUG-1996; 96WO-GB001883.  
 XX 03-AUG-1995; 95GB-00015941.  
 XX (ZENE ) ZENECA LTD.  
 PA Jenson I;  
 PI WPI; 1997-154273/14.  
 XX Chemically inducible cassette for expressing herbicide resistance gene in  
 PT plants - and derived plants, partic. for resistance to glyphosate, avoids  
 PT constitutive expression and minimises development of herbicide tolerant  
 PT weeds.

Example 3; Fig 8; 59pp; English.

CC A chemically inducible plant gene expression cassette which comprises an  
 CC inducible promoter linked to a gene (I) that confers resistance to a  
 CC herbicide, is claimed. (I) imparts resistance to N-phosphonomethyl-  
 CC glycine (glyphosate) or its salts, but may also be a gene for resistance  
 CC to, e.g. chloroacetanilides, glufosinate, sulphonyl ureas, imidazolones,  
 CC etc. The inducible promoter (e.g. alca, alcr, alda or other alcr-induced  
 CC gene promoter) is operatively linked to an alcr regulator sequence.  
 CC Induced expression of (I) avoids the risk that constitutive expression  
 CC may interfere with plant development; allows volunteer plants to be  
 CC controlled by herbicide applied without inducer and minimises the chance  
 CC that herbicide-resistant strains of weeds will arise. The expression  
 CC cassette is strictly controlled and suitable for general use in plants  
 CC (both mono- and dicotyledons). The alca/alcr gene switch was exemplified  
 CC with genes conferring resistance to glyphosate. The switch was used to  
 CC drive inducible expression of glyphosate oxidase (GOX) in plants.  
 CC Switchable GOX was expressed alone or in conjunction with constitutive  
 CC expression of 5-enol-pyruvylshikimate 3-phosphate (EPSPS) CP4. Constructs  
 CC were optimised for expression in mono- and dicotyledonous crop species.  
 CC The present sequence is that of plasmid pWJBI, based on pIBT211  
 CC containing the CamV 35 promoter with duplicated enhancer linked to the  
 CC tobacco mosaic virus translational enhancer sequence replacing the  
 CC tobacco etch virus 5' non-translated leader, and terminated with the  
 CC nopaline synthase poly(A) signal (nos). pWJBI was used in construction of  
 CC dicot vectors

SQ Sequence 1138 BP; 370 A; 253 C; 234 G; 281 T; 0 U; 0 Other;

Query Match 46.5%; Score 60.4; DB 2; Length 1138;  
 Best Local Similarity 85.8%; Pred. No. 3.3e-05;  
 Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAGACACGTGTGTCATTCTCAACAATTACCAACAAC 71  
 DB 743 AGGAAGTTCATTTCATTTGGAGAGG--ACCTCGAGTATTTTACAAACAATTACCAACAAC 800  
 QY 72 ACAAACAACAACAACAATTATCAATTTACTATTTTACAATTATCATC 117  
 DB 801 ACAAACAACAACAACAACAT--TACAATTACTATTTACAATTATCACC 844

RESULT 8  
 AAT59713  
 ID. AAT59713 standard; DNA; 1138 BP.  
 XX  
 AC AAT59713;  
 XX

DT 17-OCT-2003 (revised)  
 DT 12-MAY-1997 (first entry)  
 XX Plasmid pWJBI used for use in delta-endotoxin expression.  
 DE Gene expression cassette; promoter; alcr regulator; insecticide;  
 XX CryIA(c); CryII; crystal protein; delta-endotoxin; Bacillus thuringiensis;  
 KW Lepidoptera; insect resistance; transgenic plant; pWJBI; crop protection;  
 KW biological control; alcr; gene switch; ds.  
 XX Brassica oleracea var. botrytis; mosaic virus.  
 OS Nicotiana tabacum; mosaic virus.  
 OS Agrobacterium tumefaciens.  
 OS Chimeric.  
 XX Key Location/Qualifiers  
 FT TATA\_signal 737..743  
 FT /\*tag= c

WO9706268-A2.

20-FEB-1997.

29-JUL-1996; 96WO-GB001846.

08-AUG-1995; 95GB-00016241.

(ZENE ) ZENECA LTD.

Jenson I, Paine JAM;

WPI; 1997-154272/14.

XX Chemically inducible expression cassette - contains inducible promoter  
 PT activated by alcr regulator in presence of alcohol or ketone inducer,  
 PT used for insecticide production in plants.

Example 4; Fig 11; 52pp; English.

CC Plasmid pWJBI (AAT59713) is based on pIBT211 and contains the cauliflower  
 CC mosaic virus 35S promoter with duplicated enhancer linked to the tobacco  
 CC mosaic virus (TMV) translational enhancer sequence and a nopaline-  
 CC synthase 3' gene region. Incorporation of an endotoxin CryV (see also  
 CC AAT59702) or CryIA(c) (see also AAT59712) gene into the plasmid  
 CC downstream of the TMV enhancer provides vectors for the constitutive or  
 CC inducible expression of CryV or CryIA(c). Transformation of plants  
 CC allows the generation of transgenic plants resistant to insect pests.  
 CC (Updated on 17-OCT-2003 to standardise OS field)

SQ Sequence 1138 BP; 369 A; 254 C; 234 G; 281 T; 0 U; 0 Other;

Query Match 46.5%; Score 60.4; DB 2; Length 1138;  
 Best Local Similarity 85.8%; Pred. No. 3.3e-05;  
 Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;

QY 12 AGGAAGTTCATTTCATTTGGAGACACGTGTGTCATTCTCAACAATTACCAACAAC 71  
 DB 743 AGGAAGTTCATTTCATTTGGAGAGG--ACCTCGAGTATTTTACAAACAATTACCAACAAC 800  
 QY 72 ACAAACAACAACAACAATTATCAATTTACTATTTTACAATTATCATC 117  
 DB 801 ACAAACAACAACAACAACAT--TACAATTACTATTTACAATTATCACC 844

RESULT 9  
 AAX78852  
 ID AAX78852 standard; cDNA; 1138 BP.  
 XX  
 AC AAX78852;  
 XX 07-SEP-1999 (first entry)  
 DT  
 XX Plasmid pWJBI EcoRI-HindIII cDNA fragment.

```

XX Paraquat; antibody; light chain; herbicide; resistant; crop plant;
KW weed control; tolerant; diquat; photosynthesis inhibitor; photosystem I;
KW free radical; lipid peroxidation; electron transport; photosystem II;
KW vacuole; cell surface; cytotoxic; sensitive; ss.
XX
OS Synthetic.
XX
XX WO9932630-A1.
XX
XX 01-JUL-1999.
XX
XX 15-DEC-1998; 98WO-GB003760.
XX
XX 19-DEC-1997; 97GB-00026955.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Holt DC, Jones PG;
XX
XX WPI; 1999-405173/34.
XX
XX Herbicide binding proteins and related polynucleotides.
XX
XX Disclosure; Page 46-47; 60pp; English.
XX
XX This invention describes a novel herbicide binding protein which can
CC confer herbicide resistance activity. Crop plants, such as soybean,
CC cotton, tobacco, sugarbeet, oilseed rape, canola, flax, sunflower,
CC potato, tomato, alfalfa, lettuce, maize, wheat, sorghum, rye, bananas,
CC barley, oat, turf grass, forage grass, sugar cane, pea, field bean, rice,
CC pine, poplar, apple, grape, citrus or nut plants, transformed with a
CC herbicide binding protein gene are resistant to the herbicide. Hence,
CC weeds can be selectively controlled in a field of the transformed crops.
CC The plants are substantially resistant or tolerant to herbicides, such as
CC paraquat or diquat, that inhibit photosynthesis by accepting electrons
CC from photosystem I thus generating free radicals which cause lipid
CC peroxidation or by blocking electron transport in photosystem II. The
CC herbicide binding proteins advantageously sequester the herbicide, e.g.
CC at the cell surface or in the vacuoles of a treated plant. Sequestration
CC at the cell surface prevents the entry of the herbicide into the cell so
CC that the herbicide cannot reach its intracellular target and exert any
CC significant cytotoxic effect. The herbicide binding protein inhibits the
CC mobility of the herbicide from the application site to the whole plant
CC preventing the herbicide reaching particularly sensitive organs.
CC Additionally, tolerant plants can be produced against herbicides that
CC have more than one target site
XX
XX Sequence 1138 BP; 370 A; 255 C; 233 G; 280 T; 0 U; 0 Other;
XX
XX Query Match 46.5%; Score 60.4; DB 2; Length 1138;
XX Best Local Similarity 85.8%; Pred. No. 3.3e-05;
XX Matches 91; Conservative 0; Mismatches 11; Indels 4; Gaps 2;
XX
XX 12 AGGAAGTTCATTTCATTTGGAAATGGACAGTGTTGTCATTCTCACAAATTACCAAC 71
XX
XX 743 AGGAAGTTCATTTCATTTGGAGAGG--ACCTCAGATTTTTCACAAATTACCAAC 800
XX
XX 72 AACAAACAACAACAACATTATACAAATTACTATTACAAATTACATC 117
XX
XX 801 AACAAACAACAACAACAT--TACAAATTACTATTACAAATTACACC 844
XX
XX
XX RESULT 10
XX AAD44423
XX
XX AAD44423 standard; DNA; 157 BP.
XX
XX AAD44423;
XX
XX 29-AUG-2003 (revised)
XX
XX 13-DEC-2002 (first entry)
XX
XX pGL482 construct comprising 35S-PCISV promoter DNA.

```

```

XX Cauliflower mosaic virus 35S RNA; 35S; promoter; transgenic plant; OPR;
KW peanut chlorotic streak caulimovirus full length transcript promoter;
KW PCISV; 12-oxophytodienoic acid-100, 11-reductase gene promoter; SAG12;
KW transgenic; CaMV; chimeric; ds.
XX
XX Peanut; chlorotic streak caulimovirus.
OS
OS Cauliflower mosaic virus.
OS Chimeric.
XX
XX US6388170-B1.
XX
XX 14-MAY-2002.
XX
XX 07-APR-2000; 2000US-00545244.
XX
XX 07-APR-2000; 2000US-00545244.
XX
XX (KENT ) UNIV KENTUCKY RES FOUND.
XX
XX Gan S, Xie M, He Y;
XX
XX WPI; 2002-442888/47.
XX
XX Bi-directional promoter with common promoter linked in opposite
PT orientation to minimal promoter, useful to direct expression of
PT polynucleotide which confers agronomically significant trait to
PT transgenic plant.
XX
XX Disclosure; Col 13-16; 16pp; English.
XX
XX The invention relates to an artificial nucleic acid construct comprising
CC a bi-directional promoter having a minimal promoter operably linked in
CC opposite orientation 5' to a common promoter, where each promoter is
CC either cauliflower mosaic virus (CaMV) 35S RNA promoter (35S), peanut
CC chlorotic streak caulimovirus full length transcript promoter (PCISV),
CC Arabidopsis 12-oxophytodienoic acid-100, 11-reductase gene promoter (OPR)
CC or SAG12, and where each end of the bi-directional promoter is operably
CC linked to a polynucleotide encoding a polypeptide. The construct is used
CC to produce a transgenic plant which has an agronomically-significant
CC trait. The present sequence is pGL482 construct comprising 35S-PCISV
CC promoter DNA. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 157 BP; 57 A; 38 C; 22 G; 40 T; 0 U; 0 Other;
XX
XX Query Match 45.8%; Score 59.6; DB 6; Length 157;
XX Best Local Similarity 87.3%; Pred. No. 4.4e-05;
XX Matches 89; Conservative 0; Mismatches 9; Indels 4; Gaps 2;
XX
XX 12 AGGAAGTTCATTTCATTTGGAAATGGACA--CGTGTGTCATTCTCACAAATTACCAACA 69
XX
XX 41 AGGAAGTTCATTTCATTTGGAGAGGACAGGCTCTTGAGATCTTCAACAATTACCAACA 100
XX
XX 70 ACAACAACAACAACAACATTATACAAATTACTATTACAT 111
XX
XX 101 ACAACAACAACAACAACAT--TACAATTACTATTACAT 140
XX
XX
XX RESULT 11
XX ABL58982
XX
XX ABL58982 standard; DNA; 2200 BP.
XX
XX ABL58982;
XX
XX 22-JUL-2002 (first entry)
XX
XX HPV16-L1 encoding DNA 2.
XX
XX Human papilloma virus; HPV; HPV16-L1; HPV16-L2; vaccine; cervical cancer;
KW HPV16-L1; cytostatic; virucide; gene; ds.
XX
XX Human papillomavirus.
OS
OS Synthetic.

```

XX FH Key Location/Qualifiers  
 FT promoter 25..562  
 FT /\*tag= a  
 FT /label= 35S\_promoter  
 FT misc\_feature 565..627  
 FT /\*tag= b  
 FT /label= TMV-Overdrive-Enhancer  
 FT CDS 628..2145  
 FT /\*tag= c  
 FT /product= "HPV16-L1"  
 XX WO200238769-A2.  
 XX 16-MAY-2002.  
 XX 19-SEP-2001; 2001WO-DE003618.  
 XX 09-NOV-2000; 2000DE-01055545.  
 XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 XX (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.  
 XX Mueller M, Leder C, Kleinschmidt J, Sonnnewald U, Biemelt S;  
 XX WPI; 2002-426950/45.  
 XX P-PSDB; ABB77479.  
 XX New DNA sequences encoding human papilloma virus L1 or L2 protein, useful  
 XX in vaccines, are optimized for high-level expression in eukaryotic cells.  
 XX Claim 3; Fig 13; 39pp; German.  
 XX The invention relates to DNA sequences (I) that encode human papilloma  
 XX virus (HPV) 16 L1 protein (ABL58976, ABL58977) or L2 protein (ABL58978)  
 XX or proteins with the biological activity of L1 and L2. Expression vectors  
 XX containing (I) or a similar sequence encoding an L1/E7 fusion protein  
 XX (ABL58979-ABL58981) and the proteins encoded by them (ABB77478-ABB77483),  
 XX are useful in vaccines, especially for control of cervical cancer. (I)  
 XX are also useful for recombinant production of L1 and L2 proteins. (I) are  
 XX optimised for codon usage in eukaryotic cells and provide high yields of  
 XX L1/L2 or their fusions, without the use of viral vectors  
 XX SQ Sequence 2200 BP; 576 A; 703 C; 574 G; 347 T; 0 U; 0 Other;  
 Query Match 44.8%; Score 58.2; DB 6; Length 2200;  
 Best Local Similarity 86.4%; Pred. No. 0.00012;  
 Matches 89; Conservative 0; Mismatches 8; Indels 6; Gaps 2;  
 QY 12 AGGAAGTTCATTTCATTTGGAATGGACACCGTGTGTCATTCTCAACAATTACCAAC 71  
 DB 530 AGGAAGTTCATTTCATTTGGAATGGACACCGTGTGTCATTCTCAACAATTACCAAC 585  
 QY 72 AACAAACAACAACAACAATTATCAATTACTATTACAAATTAC 114  
 DB 586 AACAAACAACAACAACAATTATCAATTACTATTACAAATTAC 626  
 RESULT 12  
 AAQ74695  
 ID AAQ74695 standard; DNA; 5033 BP.  
 XX AAQ74695;  
 XX 25-MAR-2003 (revised)  
 DT 09-MAY-1995 (first entry)  
 XX pH11406 vector contg. the MCDV coat protein 3 gene.  
 XX MCDV; maize chlorotic dwarf virus; coat protein; plasmid vector;  
 XX plant resistance; viral infection resistance; maize plant;  
 XX maize dwarf mosaic virus; MDMV; ds.  
 XX

OS Synthetic.  
 XX Key Location/Qualifiers  
 FT CDS 119..672  
 FT /\*tag= a  
 FT /note= "Zea mays alcohol dehydrogenase 1, intron 1  
 FT fragment"  
 XX WO9421796-A2.  
 XX 29-SEP-1994.  
 XX 22-MAR-1994; 94WO-US003028.  
 XX 24-MAR-1993; 93US-00038768.  
 XX (PION-) PIONEER HI-BRED INT INC.  
 XX (USDA ) US SEC OF AGRIC.  
 XX Roth BA, Townsend R, McMullen MD;  
 XX WPI; 1994-317016/39.  
 XX DNA encoding maize chlorotic dwarf virus proteins - used to provide  
 XX plants with resistance to the virus and related viral infections.  
 XX Claim 2; Page 14-17; 40pp; English.  
 XX This sequence shows the nucleotide sequence for the plasmid designated  
 XX pPH11406. The plasmid vector contains the gene for MCDV (maize chlorotic  
 XX dwarf virus) coat protein 3 placed under control of tandem cauliflower  
 XX mosaic virus 35S promoters isolated from the 1841 strain of the virus,  
 XX and a polyA signal sequence obtained from the potato proteinase inhibitor  
 XX II (Pin II) gene that exhibits enhancer-like activity. The chimeric gene  
 XX also includes a 79 bp sequence omega' from the 5' leader region of  
 XX tobacco mosaic virus that functions as a translational enhancer; and a  
 XX Zea mays alcohol dehydrogenase 1, intron 1 fragment spanning nucleotides  
 XX 119-672, trimmed to 557 bp with Bal 31 nuclease, which has been shown to  
 XX function as an enhancer of gene expression in monocots. The DNA is used  
 XX for imparting resistance to MCDV or viruses to which MCDV infection or  
 XX resistance provides cross-resistance, including maize dwarf mosaic virus  
 XX strain A. Any or all of the three coat protein genes from MCDV can be  
 XX used to provide protection for plants. MCDV has a single, long RNA core  
 XX which is shown in AAQ74694. (Updated on 25-MAR-2003 to correct PN field.)  
 XX SQ Sequence 5033 BP; 1345 A; 1177 C; 1177 G; 1334 T; 0 U; 0 Other;  
 Query Match 43.2%; Score 56.2; DB 2; Length 5033;  
 Best Local Similarity 83.2%; Pred. No. 0.00037;  
 Matches 89; Conservative 0; Mismatches 13; Indels 5; Gaps 2;  
 QY 12 AGGAAGTTCATTTCATTTGGAATGGAC---ACGTGTGTGTCATTCTCAACAATTACCAAC 68  
 DB 1165 AGGAAGTTCATTTCATTTGGAATGGACGAGCTGCGCTTATTTTACAAATTACCAAC 1224  
 QY 69 AACAAACAACAACAACAATTATCAATTACTATTACAAATTAC 115  
 DB 1225 AACAAACAACAACAACAATTATCAATTACTATTACAAATTAC 1269  
 RESULT 13  
 AAQ95155  
 ID AAQ95155 standard; DNA; 6395 BP.  
 XX AAQ95155;  
 XX 17-APR-1996 (first entry)  
 XX Tobacco mosaic virus.  
 XX Truncated; tobacco mosaic virus; TMV; movement protein; coat protein;  
 XX frame shift; mutation; in-frame stop codon; antigen; mammal; primer;  
 XX viral vaccine; contraceptive; immune response; plant virus; insect;  
 XX



DE cDNA sequence of the genome of tobacco mosaic virus-U1.  
 XX Plant phenotype; gene trait; Nicotiana; Oryza sativa; Zea mays; Brassica;  
 KW Gossypium; Triticum; Arabidopsis; Petunia; herbicide; transgenic plant;  
 KW tobacco necrosis virus; TMV; tobacco mosaic virus; TMV; helper virus; ss.  
 XX  
 OS Tobacco mosaic virus.

XX WO200063397-A2.  
 XX 26-OCT-2000.  
 XX 17-APR-2000; 2000WO-EP003521.  
 XX 20-APR-1999; 99US-00294022.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 XX Meulewaeter F, Cornelissen M, Jacobs J, Van Eldik G, Metzlaiff M;  
 XX WPI; 2000-687182/67.

XX Identifying and isolating genes involved in determining the trait or  
 PT phenotype of plant species, by infecting plants with gene silencing  
 PT constructs targeted to the gene, and identifying plants with altered  
 PT traits.

XX Example 1; Page 50-52; 64pp; English.

XX The specification describes a method for isolating genes that determine a  
 CC trait or phenotype of a plant species. The method comprises identifying a  
 CC set of nucleic acids of genes correlated with the trait, creating a  
 CC library of gene silencing constructs in a viral RNA vector, targeting the  
 CC gene silencing constructs to the nucleic acid set, infecting a collection  
 CC of individual plants with these, identifying plants with altered traits  
 CC or phenotype, and isolating genes of the invention. The method is useful  
 CC for isolating genes involved in the determination of trait or a phenotype  
 CC of a plant such as Nicotiana, Oryza sativa, Zea mays, Brassica, Gossypium,  
 CC Triticum, Arabidopsis or Petunia. The method is also useful for  
 CC modulating the expression of selected nucleic acid sequences and for  
 CC validating the function of a nucleic acid sequence whose expression is  
 CC correlated with the presence or absence of a specific trait in plants,  
 CC but with otherwise unknown function. The method is also useful for  
 CC developing agronomically useful products such herbicides or transgenic  
 CC plants. The present sequence represents the cDNA sequence of the genome  
 CC of tobacco mosaic virus (TMV). The sequence encodes a movement protein,  
 CC and was used to construct a plasmid vector for the synthesis of an  
 CC infective hybrid tobacco mosaic virus (TMV)/ tobacco necrosis virus (TMV)  
 CC helper virus RNA. This helper virus is used in the method of the  
 CC invention

XX SQ Sequence 6395 BP; 1862 A; 1223 C; 1545 G; 1765 T; 0 U; 0 Other;

Query Match 40.0%; Score 52; DB 3; Length 6395;  
 Best Local Similarity 85.4%; Pred. No. 0.0037;  
 Matches 70; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY	46	GTCAATTTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAATTATCTACTATT	105
DB	1	GTATTTTACCAATTTACCAACAACAACAACAACAACAACAACAACAATTATCTACTATT	58
QY	106	TACAATTATCATCTAGATAACA	127
DB	59	TACAATTATCATCTAGATAACA	80

Search completed: July 3, 2005, 02:38:07  
 Job time : 202.647 secs

**This Page Blank (usp20)**



1	130	100.0	452	4	US-09-263-692A-1	Sequence 1, Appl
2	89	68.5	89	4	US-09-263-692A-13	Sequence 13, Appl
3	60.4	46.5	1138	3	US-09-011-151-8	Sequence 8, Appl
4	60.4	46.5	1138	3	US-09-011-151-9	Sequence 9, Appl
5	59.6	45.8	157	3	US-09-545-244A-9	Sequence 9, Appl
6	56.2	43.2	5033	1	US-08-038-768A-1	Sequence 1, Appl
7	52	40.0	6395	2	US-08-687-559-2	Sequence 2, Appl
8	52	40.0	6395	2	US-09-259-741-1	Sequence 1, Appl
9	52	40.0	6395	3	US-09-037-751-1	Sequence 1, Appl
10	52	40.0	6395	3	US-09-466-422-1	Sequence 1, Appl
11	52	40.0	6395	4	US-09-401-415-2	Sequence 2, Appl
12	52	40.0	6395	4	US-09-962-527-1	Sequence 2, Appl
13	52	40.0	6425	3	US-09-259-741-3	Sequence 3, Appl
14	52	40.0	6425	3	US-09-037-751-3	Sequence 3, Appl
15	52	40.0	6425	3	US-09-466-422-3	Sequence 3, Appl
16	52	40.0	6425	4	US-09-962-527-3	Sequence 3, Appl
17	52	40.0	6439	3	US-09-259-741-2	Sequence 2, Appl
18	52	40.0	6439	3	US-09-037-751-2	Sequence 2, Appl
19	52	40.0	6439	3	US-09-466-422-2	Sequence 2, Appl
20	52	40.0	6439	4	US-09-962-527-2	Sequence 2, Appl
21	52	40.0	6446	3	US-09-259-741-5	Sequence 5, Appl
22	52	40.0	6446	3	US-09-037-751-5	Sequence 5, Appl
23	52	40.0	6446	3	US-09-466-422-5	Sequence 5, Appl
24	52	40.0	6446	4	US-09-962-527-5	Sequence 5, Appl
25	52	40.0	6475	3	US-09-259-741-4	Sequence 4, Appl
26	52	40.0	6475	3	US-09-037-751-4	Sequence 4, Appl
27	52	40.0	6475	3	US-09-466-422-4	Sequence 4, Appl







RESULT 12  
US-09-962-527-1  
; Sequence 1, Application US/09962527  
; Patent No. 6740740  
; GENERAL INFORMATION:  
; APPLICANT: GARGER, STEPHEN  
; HOLTZ, R. BARRY  
; MCCULLOCH, MICHAEL  
;

RESULT 13  
US-09-259-741-3  
Sequence 3, Application US/09259741  
Patent No. 6033895  
GENERAL INFORMATION:  
APPLICANT: GARGER, STEPHEN  
APPLICANT: HOLTZ, R. BARRY  
APPLICANT: MCCULLOCH, MICHAEL  
APPLICANT: TURPEN, THOMAS  
TITLE OF INVENTION: A PROCESS FOR  
TITLE OF INVENTION: PURIFYING V  
TITLE OF INVENTION: SOURCES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howrey & Simon  
STREET: 1299 Pennsylvania Ave  
CITY: Washington  
STATE: DC

```
;
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/259,741
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; US-09-259-741-3
;
; Query Match 40.0%; Score 52; DB 3; Length 6425;
; Best Local Similarity 64.6%; Pred. No. 0.0002;
; Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;
;
Qy 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAUAUCCAAACAACAACAACAACAACAACAUA--UACAAUUUUAUUU 58
;
Qy 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAAUUACAUAUGGCAUACACA 80
;
;
; RESULT 14
; US-09-037-751-3
; Sequence 3, Application US/09037751
; Patent No. 6037456
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: .USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,751
; FILING DATE: 10-MAR-1998
; CLASSIFICATION:
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; US-09-037-751-3
;
; Query Match 40.0%; Score 52; DB 3; Length 6425;
; Best Local Similarity 64.6%; Pred. No. 0.0002;
; Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;
;
Qy 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAATTATCAATTACTATT 105
Db 1 GUUUUUUACAUAUCCAAACAACAACAACAACAACAACAUA--UACAAUUUUAUUU 58
;
Qy 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAAUUACAUAUGGCAUACACA 80
;
;
; RESULT 15
; US-09-466-422-3
; Sequence 3, Application US/09466422
; Patent No. 6303779
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: MCCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/466,422
; FILING DATE: 17-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
```



**This Page Blank (usp...)**



Result No.	Score	Query Match	Length	DB	ID	Description
1	63.6	48.9	2186	21	US-10-784-295-13	Sequence 13, Appl
2	54.8	42.2	368	9	US-09-331-631A-29	Sequence 29, Appl
3	54.8	42.2	368	16	US-10-147-095-29	Sequence 29, Appl
4	52	40.0	6395	10	US-09-962-527-1	Sequence 1, Appli
5	52	40.0	6395	17	US-10-338-592-2	Sequence 2, Appli
6	52	40.0	6395	19	US-10-828-029-1	Sequence 1, Appli
7	52	40.0	6425	10	US-09-962-527-3	Sequence 3, Appli

```

Db      312 AGGAAGTTCAATTCATTTGGAGAGACAGGCTTCTTGAGATCTCTTCAACAATTACCAACA 377
Qy      70 ACAACAACAACAACAACAACATTATATACAATTACTATTACAAATTACA 115
Db      372 ACAACAACAACAACAACAACAT--TACATTACTATTACAAATTACA 415

RESULT 2
US-09-331-631A-29
; Sequence 29, Application US/09131631A
; Patent No. US20020168392A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001APC
; CURRENT APPLICATION NUMBER: US/09/331.631A
; CURRENT FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A synthetic nucleotide sequence which can be used
; OTHER INFORMATION: for the expression and secretion of MIAMP2c.
; OTHER INFORMATION: containing the leader sequence from SEQ ID NO:11
; OTHER INFORMATION: and SEQ ID NO:5.
; NAME/KEY: CDS
; LOCATION: (103)...(333)
US-09-331-631A-29

Query Match 42.2%; Score 54.8; DB 9; Length 368;
Best Local Similarity 76.9%; Pred.No. 0.00064;
Matches 80; Conservative 0; Mismatches 22; Indels 2; Gaps 1

Qy      27 TTTGGAAATCGACACGCTGTGTTCATTTCACAACAATTACCAACAACAACAACAACAACA 86
Db      4 TCTAGAGCGCGCGCTGCAGCTATTTTACACATTTACCAACAACAACAACAACAACAACA 63

Qy      87 ACATTAACAATTACTATTATTACAATTACTTAGATAAACAATG 130
Db      64 ACAT--TACAATTACTATTATTAACAATTACAGGATCCACAACAATG 105

RESULT 3
US-10-147-095-29
; Sequence 29, Application US/10147095
; Publication No. US20030171274A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie L.
; TITLE OF INVENTION: ANTIMICROBIAL PROTEINS
; FILE REFERENCE: CULN23.001APC
; CURRENT APPLICATION NUMBER: US/10/147,095
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US/09/331.631A
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: PCT/AU97/00874
; PRIOR FILING DATE: 1997-12-22
; PRIOR APPLICATION NUMBER: AU PO 4275
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0

```

```
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-962-527-1

Query Match          40.0%; Score 52; DB 10; Length 6395;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GUUUUUUACAACAUAUACCAACAACAACAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 UACAUAUUAUACAUGGCAUACACA 80

RESULT 5
US-10-338-592-2
; Sequence 2, Application US/10338592
; Publication No. US20030208792A1
; GENERAL INFORMATION:
; APPLICANT: Fitch, John H.
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; FILE REFERENCES: TSRI 312.1C1
; CURRENT APPLICATION NUMBER: US/10/338,592
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 09/401,415
; PRIOR FILING DATE: 1999-09-21
; PRIOR APPLICATION NUMBER: US 08/687,559
; PRIOR FILING DATE: 1996-11-18
; PRIOR APPLICATION NUMBER: PCT/US95/01467
; PRIOR FILING DATE: 1995-02-03
; PRIOR APPLICATION NUMBER: US 08/192,477
; PRIOR FILING DATE: 1994-02-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6395
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus
US-10-338-592-2

Query Match          40.0%; Score 52; DB 17; Length 6395;
Best Local Similarity 85.4%; Pred. No. 0.007;
Matches 70; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTATTTTACAACAATTACCAACAACAACAACAACAACAACAACAACAACAAT--TACAATTACTATT 58

QY 106 TACAATTACATCTAGATAAACA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 TACAATTACAATGGCATAACA 80

RESULT 6
US-10-828-029-1
; Sequence 1, Application US/10828029
; Publication No. US20040171813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/828,029
; FILING DATE: 20-Apr-2004
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-828-029-1

Query Match          40.0%; Score 52; DB 19; Length 6395;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

QY 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAACAACAATTATCAATTACTATT 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GUUUUUUACAACAUAUACCAACAACAACAACAACAACAACAACAACAACAUA--UACAUAUUAUUU 58

QY 106 TACAATTACATCTAGATAAACA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 UACAUAUUAUACAUGGCAUACACA 80

RESULT 7
US-09-962-527-3
; Sequence 3, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
;
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962.527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-962-527-3
Query Match 40.0%; Score 52; DB 10; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

Qy 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATTAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUCCAAACAACAACAACAACAACAACAACAU--UACAUAUUCUAUU 58

Qy 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUACAUAUGGCAUACACA 80

RESULT 8
US-10-828-029-3
; Sequence 3, Application US/10828029
; Publication No. US20040171813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/828,029
; FILING DATE: 20-Apr-2004
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; APPLICATION NUMBER: 09/037,751

Qy 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATTAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUCCAAACAACAACAACAACAACAACAACAU--UACAUAUUCUAUU 58

Qy 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUACAUAUGGCAUACACA 80

RESULT 9
US-09-962-527-2
; Sequence 2, Application US/09962527
; Publication No. US20030049813A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCCULLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,527
; FILING DATE: 24-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: 10-march-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 00801.0140.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-828-029-3
Query Match 40.0%; Score 52; DB 19; Length 6425;
Best Local Similarity 64.6%; Pred. No. 0.007;
Matches 53; Conservative 17; Mismatches 10; Indels 2; Gaps 1;

Qy 46 GTCATTCTCAACAATTACCAACAACAACAACAACAACAACAACAATTATTAATTACTATT 105
Db 1 GUUUUUUACAACAAUUUCCAAACAACAACAACAACAACAACAACAU--UACAUAUUCUAUU 58

Qy 106 TACAATTACATCTAGATAAACA 127
Db 59 UACAUAUACAUAUGGCAUACACA 80
```







This page blank (USP12)